Differentiated Services framework

Risto Mononen

Nokia Telecommunications

P.O. Box 300, FIN-00045 NOKIA GROUP

Risto.Mononen@nokia.com

Abstract

Voice over IP and other jitter sensitive applications have raised the demand for Quality of Service guarantees in the Internet. The traffic must be classified and prioritized according to application specific needs. The Differentiated Services framework provides means for stateless low-cost classification. This paper summarizes the IETF's DiffServ Internet Draft.

Introduction

Differentiated Services (DiffServ, DS) has background in the IETF's Integrated Services (IntServ) and Resource Reservation Protocol (RSVP) work. Lots of expectations were laid on IntServ. However, IntServ nodes have turned out to keep too much per-flow state information to be scalable to large networks.

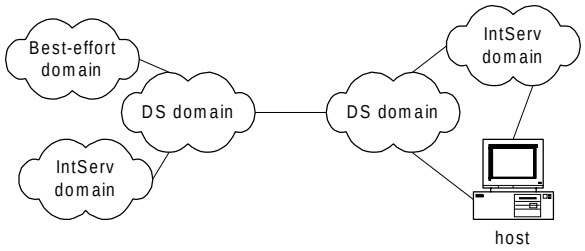

Figure 1: DiffServ deployment

DiffServ must interoperate with current IntServ and traditional best-effort networks. Hosts may attach to any of these domains. DiffServ aggregates Quality of Service (QoS) information from several flows, tries to push complexity to DS domain boundaries and works on per-packet basis as much as possible. Aggregation means DS field (IPv4 TOS or IPv6 Traffic Class) based packet classification and better scaling in comparison with IntServ. More stateful approaches, like multifield classification and RSVP signaled QoS parameters, have also been proposed but for the start the DiffServ standardization seems to prefer simple stateless forwarding paradigm as far as possible. This paper does not consider these more complex solutions for the following reasons:

1. The stateless forwarding with classes can be taken into use with minimal effort which makes it viable in the near future.

2. There already exist several resource reservation (RSVP etc.) and virtual circuit (ATM, MPLS, etc.) oriented technologies. Dynamic QoS and statefulness in DiffServ partly overlaps with these known technologies and does not actually introduce anything new into the field.

Inside each DS class the traffic is still best effort in the DS domain. Policing and pricing prevents all senders from requesting the highest priority, which would degenerate the DS domain into traditional best-effort network without real QoS support. Policing takes place at the network boundaries. It adds complexity at the boundaries but is necessary to protect the backbone from overloading.

Differentiated Services framework [2] and DiffServ charter [1] have been the main sources for this paper. Both contain lots of useful links for the interested reader. Most of the references are still draft phase Internet documents due to DiffServ concept novelty.

The framework Internet Draft considers end-to-end and network level issues of Quality of Service. The draft outlines possible contents of agreements between the Internet service provider and a customer. Although mostly pricing oriented, part of the agreement is assumed to cover some more technical issues like traffic control. Differentiated Services architecture RFC, another Internet document, describes network element and hop-by-hop behavior. The architecture part is somewhat more technical than the framework. Both are needed to fully understand what are the goals of Differentiated Services and how it is implemented.

1. End-to-end Services

The end-user applications define the required service level that should be guaranteed end-to-end to fulfill the customer's needs. Per-hop-behavior (PHB) is the individual router's contribution to overall QoS. Concatenation of PHBs must match suitably to get the correct end-to-end behavior. The customer – provider agreements express the Service Level in such a way that the contracting parties can police traffic at the edge to avoid DS domain congestion.

1.1 Per-hop-behavior

Per-hop-behavior (PHB) is the basic DiffServ building block. The DiffServ architecture [8] and PHB specific documents, e.g. Assured forwarding [4] and Expedited forwarding [5], define PHBs and codepoints more closely.

Assured forwarding (AF) divides the packets to four priority classes. The classes are further divided into three drop precedences. Inside a class packet ordering is preserved.

Table 1: Assured forwarding codepoints

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

PHB definition is still ongoing work in the DiffServ working group. Internet draft [6] defines how the PHBs should be defined. The framework draft mentions the above two PHBs; after the draft at least two "colored marker" drafts [11, 12] have been published.

The PHBs must match all the way from sending host to the receiver in order to give the customer real benefit. Therefore the DS field codepoints should have uniform standardized meaning in all DS devices. Technically it is possible that the traffic passes several DS domains between the endpoints. Different policies and lack of agreements between providers are more serious obstacles for large scale DiffServ deployment than any strictly technical issue.

1.2 Service level specification

A Service Level Agreement (SLA) between the customer and provider contains payment terms and other mostly non-technical items. The technical part of the agreement is called Service Level Specification (SLS). Its contents describe overall features, QoS and performance.

Boundary conditioning prevents DS domain from overloading. A Traffic Conditioning Specification (TCS) defines the exact policing at the boundary. DiffServ framework [2] describes conditioning with the following simple TCS format:

> DS-Mark : Profile : Scope : Disposition of non-conforming traffic

DS-Mark is the DS field value, or combination of DS field and other header fields (multifield classifier), which are used for packet classification. Profile tells what is the maximum amount of DS domain ingress traffic the provider has committed to carry in the class in question. Service scope defines the topological extent of the TCS entry. The last disposition entry defines actions on the packets that exceed the agreed profile. Typically the boundary marks excess traffic with a lower class or drops the packets altogether.

The service scope is still partly an open question and raises discussion on the DiffServ mailing list [13]. According to the framework [2] the scope is a sender-oriented view of expected service level between an ingress point and set of egress points. The framework claims it is not practical to define receiver-oriented egress point scopes since every possible ingress point should have a corresponding traffic profile. In addition the ingress and egress SLSs might conflict. The mailing list comments and Receiver control draft [14] point out the receiver's need to control incoming traffic priorities especially on access links. Both organization's access link and a host's (low speed) last-hop are in danger of a service denial attacks if such control is not possible. The access and core networks apply different DS field semantics in the proposed solution. The access' special handling may be statically configured or signaled. Signaling in this context resembles the way RSVP and DiffServ interact at the edges between LAN and DS domain [15]. Clearly receiver control provides important features to the end-user, but it is still unclear if it justifies the added complexity.

The service level can be defined in a qualitative or quantitative way. A qualitative service is defined in relation to other services, which in turn may be either qualitative or quantitative. A quantitative service has absolute bounds, e.g.:

"90% of packets experience less than 5ms delay"

Qualitative services are simple to implement but the customer cannot verify the QoS improvement over the reference service class. Typically a qualitative service's PHB fixes only traffic priority or shaping policy at DS domain ingress. In principle the customer may try to measure both the QoS and reference traffic at DS boundary. But since the higher QoS guarantees are typically needed in a congested network, the measurements may be most of the time unable to show any improvement since congestion does not exist! In this case the customer can only trust the provider and consider whether the qualitative service is worth paying.

Quantitative services are much harder to implement in a large scale. The provider must provision carefully to fulfill the assigned strict bounds. Provisioning must also somehow add up all the quantitative SLSs in the DS domain, as explained in section Intra-domain provisioning.

1.3 Service examples

The framework draft gives examples of possible DiffServ applications. It describes one qualitative and two quantitative services.

The first example is a Better than Best-Effort (BBE) service. It is a qualitative service where the BBE class gets higher priority than normal traffic. E.g. a web content provider might use BBE service from his own ingress point to any egress point. This way the provider will get better delay or loss performance than a competing site. The performance improvement is not visible to the customer – at least not from a trivial ingress point's traffic monitoring. Example service's TCS:

> AF11 Mark : 1 Mbps : Any egress point : Excess traffic handled by marking with AF13 mark

The Assured forwarding (AF) PHB is defined in [4]. The TCS states that AF11 class traffic up to 1 Mbps is forwarded as is. Boundary conditioner increases excess packets' drop probability with AF13 mark. Packet ordering is not affected since traffic class AF1 is the same even for excess traffic. The above BBE service looks quite simple, provides real benefit to a customer and is an adequate candidate for the first DiffServ implementations.

Leased line Emulation is a quantitative service, which tries to minimize both packet delay and loss up to an agreed maximum rate. Boundary conditioner discards excess packets. Example TCS:

> EF-Mark : 100 kbps : Egress point B : Discard non-conforming traffic

Concurrent leased line emulation users in the same DiffServ domain must not exceed the capacity reserved for the traffic class. The provider must overengineer the network to carry the sum of all the users or rely on statistical multiplexing and the very likely circumstance that everyone is not sending simultaneously.

Assured Media Playback Service is quantitative like Leased line Emulation but the profile contains burstiness specification in addition to bandwidth. Compressed video or audio is its likely application. The price would be lower than that of Leased line Emulation since the traffic is more flexible in nature, i.e. the burstiness allows certain amount of statistical multiplexing gain. Example TCS:

> AF11-Mark : 100 Kbps sustained, 100 Kb bursts tolerated at up to 200 Kbps : Egress point B : Excess burst traffic over sustained rate marked with AF12-mark : Non-conforming traffic marked with AF13-mark : Max latency = 1 second

The TCS looks almost as detailed as an ATM traffic descriptor [16]. Probably the provider must convert this TCS into some "equivalent bandwidth" form to make it comparable with e.g. the leased line service requirements when provisioning the DS domain. The latency part of the TCS cannot be guaranteed without some kind of flow specific virtual circuit. Again overengineering is the best a provider can do.

Both qualitative and quantitative services in the same network might be needed if some customers want e.g. Leased line Emulation for Voice over IP (VoIP), others BBE for business applications and still some others traditional best-effort. The provider must allocate resources carefully and probably it will take some time to learn how to dimension a DS domain.

2. Congestion avoidance

Inside each DS class the traffic is still best effort as was noted in the Introduction. In the worst case all customers can send their packets tagged with the highest priority DS field, which would degenerate the DS domain into traditional best-effort network without real QoS support. To guarantee, at least statistically, higher service levels to their actual users, the provider can:

1. Price higher classes expensively.

2. Police the ingress traffic into the DS domain.

3. Provision the network with enough capacity to carry the traffic.

Pricing actually makes the QoS services and applications less attractive from the customer point of view. Besides the "feedback" from pricing comes typically after one or two month's period which gives malicious user plenty of time to disturb the other network users. Policing and provisioning are considered further below.

2.1 Boundary policing

Policing takes place at all the DS domain boundaries. In Figure 1: DiffServ deployment this means all boundaries including the two DS domains. Policing adds complexity at the boundaries but is necessary to protect the backbone from overloading and prevent certain type of denial of service attacks against the DS domain and its other users.

The DS domain provider's boundary nodes police the ingress traffic. From the provider point of view traffic control prevents overloading the DS domain. To the customer policing provides means to detect if (quantitative) service agreement has been fulfilled. The customer may also prioritize and mark his own traffic before entering DS domain to get the best QoS for the money. Here the word "customer" covers both an end user and a transit network's operator; also boundaries between consecutive domains need conditioners.

The architecture RFC [8] defines the following traffic conditioner components: meter, marker, shaper and dropper. Figure 2: DiffServ traffic conditioner shows the logical view of the conditioner. DiffServ boundary router requirements [17] and PHB specifications [4, 5, 11, 12] refine how the conditioners are used within each specific QoS class.

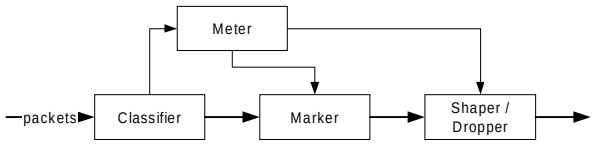

Figure 2: DiffServ traffic conditioner

Section Service level specification explained how SLS and TCS define the boundary conditioning policy. In the DiffServ starting phase these SLSs will be static. The SLS may contain temporal changes, e.g. to reflect cyclic changes in the customer needs or DS domain load patterns, but the agreement itself is still static. Dynamic SLSs change without human interaction. They require some kind agent and protocol to agree on new SLS. The dynamic SLSs should still handle aggregate traffic instead of single flows. Aggregation is another topic of discussion (like receiver-oriented service scope) which is only recommended but left open to implementations in the framework draft.

2.2 Intra-domain provisioning

DS domain dimensioning is the most fundamental way of provisioning. The ISP's goal is to fulfill the agreed service levels (i.e. SLAs) with the least equipment. Router configuration and possibly some form of signaling provide additional static and dynamic means of provisioning.

From the network element point of view provisioning means installing enough physical equipment (router and transmission capacity) and configuring their logical DiffServ behavior. The task is especially important in the boundary routers where policing and most of the complexity takes place. Also internal policers may be used in key nodes.

The combined effect of several customers sending packets into the same DS domain is hard to predict. The framework document gives some general guidelines on what should be taken into account when provisioning. It does not specify any concrete requirements or profile dependent formulas for calculating the needed capacity. DiffServ is still at its beginning and probably time and experience will bring out the best practices.

The fact that there are both quantitatively and qualitatively defined service levels sets its own bounds on provisioning. The quantitative services must always meet the TCS bandwidth and delay requirements, no matter how much quantitative traffic has entered the DS domain. Thus higher priority and lower drop precedence to quantitative services, i.e. their unique PHB codepoints, solves part of the problem. On the other hand the quantitative services must not starve altogether, which means the higher priority classes must be policed carefully at the DS domain ingress.

Configuration may contain statically managed and dynamically signaled parts. In the dynamic case the DiffServ nodes may bury even into micro-flow level details. This level of dynamics is probably needed only if dynamic SLS are taken into use. However, DiffServ framework proposes real-time traffic measurement based re-configuration, which could enhance the efficiency of implementing qualitative services. The traffic patterns of qualitative services are less predictable than those of quantitative ones, and the former may get real benefit from more dynamic configuration. Several protocols

have been mentioned as possible ways to deliver configuration data into the DS domain: SNMP, CLI, RSVP, COPS, and LDAP.

2.3 Inter-domain provisioning

The enduser's connection very likely spans several DS domains. End-to-end QoS requires sufficient resources from all the nodes on the way. This inter-domain provisioning will be much harder to achieve than internal provisioning. Technically both are about the same level issues with other provider – customer boundaries. However, getting to a verifiable satisfactory agreement between possibly competing ISPs will probably be a tedious task.

The framework document mentions another more technical problem in inter-domain provisioning. The DS domains may provide different services and differing implementations of the same service. The providers must therefore map the services at the (ingress or egress) boundary to the peer domain services. In addition even the same PHBs or their codepoints may differ, at least in principle. Current pace of PHB drafts and standard codepoints would suggest that this mismatching would not come true.

3. Deployment Scenarios

The framework document [2] describes a scenario for bringing DiffServ into an existing IP network. DiffServ deployment at the edge between customer and the provider is one thing to consider and evolution from non-DS domain to a DS domain another.

3.1 Deployment at DS domain edges

In the first phase the customer does not have any DiffServ aware equipment. An agreement with the ISP states how the provider's edge router will condition (classify, mark and shape) the ingress traffic to DS domain. Conditioning is a value-added service the ISP provides. The agreement is based on administrative knowledge of the customer's traffic.

In the second phase the customer's egress router is DiffServ capable and takes care of classification, marking and shaping. The provider just polices the incoming aggregate traffic. In this case the customer can more flexibly adapt to changes in traffic patterns since new priority scheme can be taken into use without provider's actions. This case covers all cases where the customer's router or routers are DiffServ aware and the hosts below are not.

In the third phase the customer has DiffServ equipment all the way down to individual hosts. The boundary between DS and non-DS domain lies between the application and host's protocols stack, which is usually part of the operating system. The application programs have access to IPv4 TOS bits [10] even today but the protocol stack does not prioritize the packets according to TOS field. Prioritizing the incoming packets may expose the host to denial of service attacks. The architecture [8] and DS field [7] RFCs discuss security issues further.

3.2 DS and non-DS router co-operation

DS routers can be added into a legacy network with the goal in improving the QoS the network can provide. Both the ISPs and customers can use this stepwise development. The new DS routers classify, mark and shape traffic according to some user defined policy. Probably the routers initially use a multi-field (MF) classifier, which selects packets based on several header fields of the packet, including transport layer headers.

The framework document [2] suggests that the remaining non-DS routers leave the DS field untouched. A non-DS router must forward the packets like other best-effort traffic. The last non-DS routers probably reside at points that never or rarely experience congestion. Their effect on overall QoS is thus negligible and the described stepwise deployment is quite likely alternative.

It is also possible to reply with ICMP "destination unreachable" to the sender [9]. The way of thinking here is "hard" QoS, which is essential part of packet forwarding, and must not be compromised. ICMP reply notifies the sender of lack of QoS support. Retry with cleared DS field may succeed but that is about all a sending host can do to recover. The IP QoS book [9] acknowledges that Internet's "be liberal on what you receive" philosophy prefers the former, untouched DS field approach.

4. Inter-operability With RSVP/integrated Services

Several vendors have announced RSVP support in the host protocol stack [15], access network and LAN environments. There its scalability weaknesses do not appear like in core networks. In the hosts Microsoft is dominating player and its NT 5.0 will support RSVP. The framework draft [2] describes RSVP over DiffServ and Parallel operation to make the two QoS concepts interoperate. One Internet draft [3] is completely dedicated to RSVP over DiffServ operation.

RSVP over DiffServ uses DS domain as a scalable core network for stub IntServ domains. The edge between the domains handles RSVP messages and makes appropriate admission decisions. It forwards the RSVP messages transparently over the DS network, which appears as a single RSVP hop to the IntServ network. QoS aggregation means many-to-one mapping from RSVP classes to DiffServ codepoints. Per-flow state is not needed inside DS domain even though the edge is IntServ aware.

Parallel mode operation does actually not affect DiffServ at all, at least from the network point of view. Parallel mode means that the same node has both DiffServ and IntServ capabilities and uses them independently. The model is usable in hosts with new RSVP aware applications and legacy applications without signaling. The hosts' protocol stack can mark DS codepoints on the legacy applications' behalf and provide some level of QoS. New applications use RSVP to tell exactly the required service level with possibly cheaper price.

5. Security and Tunneling Considerations

The architecture [8] and DS field [7] RFC discuss DiffServ security issues. In the architecture paper theft, denial of service attacks and tunneling is analyzed in detail. The framework document [2] mentions thread to accounting, DS domain configuration and statistical data collection. The problems are not DiffServ specific but more general in the Internet and under consideration in the IETF's Authentication, Authorization and Accounting (AAA) working group.

The packets DS field and its codepoint is the fundamental classification criterion. Multifield classifiers have also been mentioned but their significance is lesser due to larger processing load and security issues. Other than IP header based classification will become impossible when IPsec encryption becomes more common in the packet payload. Multifield classification is of course possible before the packet enters IPsec tunnel.

6. Conclusions

DiffServ has been one of the most promising Internet QoS technologies for a while. Its strength lies in the "low-end" or "soft" QoS: stateless better-than-best-effort classes can be added to a traditional IP network with minimal cost. More dynamic, stateful network and micro-flow oriented characteristics can be added up to a fully connection oriented signaled flows. These "high-end" features will provide better service levels and decrease the required overcapacity. On the other hand they bring the complexity of the system closer to SVC management and CAC in ATM networks since the problems to be solved are essentially the same. But as noted above and in section Deployment Scenarios, the evolutionary growth of DiffServ is a viable alternative. Signaling and hard QoS remains as an open option.

7. References

[1] Carpenter, et al: Differentiated Services (DiffServ), IETF Working Group charter, March 30 1999, http://www.ietf.org/html.charters/diffserv-charter.html

[2] Bernet, et al: A Framework for Differentiated Services, Internet Draft, February 1999, http://www.ietf.org/internet-drafts/draft-ietf-diffserv-framework-02.txt

[3] Bernet, Ed. Et al: Interoperation of RSVP/IntServ and DiffServ Networks, Internet Draft, February 26, 1999, http://www.ietf.org/internet-drafts/draft-ietf-diffserv-rsvp-02.txt

[4] Heinanen. Assured Forwarding PHB Group, INTERNET DRAFT, February 1999, http://www.ietf.org/internet-drafts/draft-ietf-diffserv-af-06.txt

[5] Van Jacobson et al: An Expedited Forwarding PHB, Internet Draft, February 1999, http://www.ietf.org/internet-drafts/draft-ietf-diffserv-phb-ef-02.txt

[6] Kathleen Nichols et al: Format for DiffServ Working Group Traffic Conditioner Drafts, INTERNET DRAFT, February 1999, http://www.ietf.org/internet-drafts/draft-ietf-diffserv-trafcon-format-00.txt

[7] Nichols, et al: Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers, Internet RFC 2474, December 1998, ftp://ftp.isi.edu/in-notes/rfc2474.txt

[8] Blake, et al: An Architecture for Differentiated Services, Internet Request for Comments 2475, December 1998, ftp://ftp.isi.edu/in-notes/rfc2475.txt

[9] Paul Ferguson and Geoff Huston: Quality of Service – Delivering QoS on the Internet and in Corporate Networks, Wiley & Sons, 250p, ISBN 0-471-24358-2.

[10] Wright, Gary R.; Stevens, Richard W.: TCP/IP illustrated, volume 2: the implementation, Addison-Wesley Publishing Company, Reading, Ma, 1995, 1174p, ISBN 0-201-63354-X.

[11] Heinanen & Guerin: A Single Rate Three Color Marker, INTERNET DRAFT, March, 1999, http://search.ietf.org/internet-drafts/draft-heinanen-diffserv-srtcm-00.txt

[12] Heinanen & Guerin: A Two Rate Three Color Marker, INTERNET DRAFT, March, 1999, http://search.ietf.org/internet-drafts/draft-heinanen-diffserv-trtcm-00.txt

[13] Borje Ohlman: More framework-02 comments, DiffServ mailing list, 01 Apr 1999, http://www-nrg.ee.lbl.gov/diff-serv-arch/msg03660.html

[14] Borje Ohlman, et al: Receiver control in Differentiated services, INTERNET-DRAFT, 30 September 1998, http://search.ietf.org/internet-drafts/draft-ohlman-receiver-ctrl-diff-01.txt

[15] Robertson, Joe: Quality of Service: DiffServ and ATM, ATM year 98 Europe Conference Proceedings, London, 16 September 1998

[16] Stallings, William: High-speed networks: TCP/IP and ATM design principles, Prentice-Hall, Upper Sadle River, NJ, 1998, 576p, ISBN : 013-525965-7

[17] Y. Bernet, et al: Requirements of DiffServ Boundary Routers, Internet Draft, November, 1998, http://search.ietf.org/internet-drafts/draft-bernet-diffedge-01.txt