Knieps, Günter

Working Paper
The evolution of the generalized differentiated services architecture and the changing role of the Internet engineering task force

Diskussionsbeitrag, No. 147

Provided in Cooperation with:
Institute for Transport Economics and Regional Policy, University of Freiburg

Suggested Citation: Knieps, Günter (2013) : The evolution of the generalized differentiated services architecture and the changing role of the Internet engineering task force, Diskussionsbeitrag, No. 147, Albert-Ludwigs-Universität Freiburg, Institut für Verkehrswissenschaft und Regionalpolitik, Freiburg i. Br.

This Version is available at:
http://hdl.handle.net/10419/83661

Terms of use:
Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.
The Evolution of the Generalized Differentiated Services Architecture and the Changing Role of the Internet Engineering Task Force*

by Günter Knieps

Discussion Paper
Institut für Verkehrswissenschaft und Regionalpolitik
No. 147 – August 2013

Abstract:
The changing role of the Internet Engineering Task Force (IETF) standard setting process from designing and implementing the best effort TCP/IP protocol as a universal standard towards a platform for dealing with the increasing need for variety in the design of a Quality of Service (QoS) differentiated traffic management architecture is demonstrated. The IETF’s contributions to a flexible open transmission architecture able to supply the required transmission qualities for the different applications provide the relevant pillars towards a Generalized Differentiated Service (DiffServ) architecture. Furthermore, the role of entrepreneurial traffic management within the Generalized DiffServ architecture and the division of labor between the IETF and entrepreneurial traffic management is analyzed. Within the “umbrella” architecture of Generalized DiffServ with the potential to combine basic elements of QoS differentiated traffic architectures a flexible framework for entrepreneurial traffic quality differentiation strategies is evolving. Its basic characteristic is market driven network neutrality with all applications bearing the opportunity costs of their required traffic capacities. As a consequence an artificial market split between best effort TCP and managed services would conflict with the integrated service approach of the IETF. Finally, the implementation of Generalized DiffServ via Next Generation networks is considered.

Prof. Dr. Günter Knieps, Institut für Verkehrswissenschaft und Regionalpolitik, Universität Freiburg, E-mail: guenter.knieps@vwl.uni-freiburg.de

* Paper to be presented at the 41st Research Conference on Communication, Information and Internet Policy (TPRC), September 27-29, 2013, George Mason University, Arlington, VA. Helpful comments by Volker Stocker are gratefully acknowledged.
1. Introduction

Due to the transition from narrowband access to broadband access, heterogeneous requirements for traffic qualities become increasingly important, taking into account the necessities for prioritization of data packets and quality of service guarantees. A transition to active network management and subsequently to a more “intelligent” Internet traffic architecture is required. In the meantime, the role of the Internet Engineering Task Force (IETF) has changed from developing and enforcing the best effort transmission control protocol (TCP) assigning all data packets equal priority to a platform for dealing with the increasing need for variety in the design of heterogeneous traffic management infrastructures and the continuous search for new technological solutions.

More than two decades ago the standard setting processes regarding Quality of service (QoS) differentiation were initiated, developing (de facto) standards for basic components for Integrated Services/Resource Reservation Protocol (IntServ/RSVP) and Differentiated Services (DiffServ) architectures. The analysis of this standard setting process during the last two decades shows that neither DiffServ architecture nor IntServ/RSVP has reached the final status of Internet standard. Nevertheless the proposed standards have in the meantime gained the status of a de facto standard within the Internet community.¹ Moreover, important developments of network management design for Generalized DiffServ networks, inserting flow based IntServ transmission into premium traffic classes of DiffServ architecture, did not even reach standards track status, but are considered to be on the non-standards track. From the perspective of fostering the evolutionary search for innovative traffic management architectures the proposed (de facto) standards and the increasing role of optional (non-obligatory) non-standards tracks for quality of service differentiated network management should not be considered as a weakness but as a strength of the IETF.

¹ Simcoe (2012, FN 26, p. 318) already considers the publication of a proposed standard as the de facto relevant end of the standards track, because further steps on the track (draft standards and Internet standards) do not constitute new protocols but the formal recognition of widespread implementation and deployment.
Due to the transition from narrowband to broadband Internet entrepreneurial decisions regarding the choice of traffic management investment and the required capacity allocation mechanisms for fulfilling the heterogeneous requirements for traffic qualities become increasingly important. As a standard setting agency the IETF cannot take over the required entrepreneurial decisions. In contrast to the technical neutrality of best effort TCP, the concept of market driven network neutrality becomes the relevant reference point in the context of competition policy. Under competition an entrepreneurial search for quality of service differentiation arises, avoiding incentives for Internet traffic service providers to discriminate between possible network applications on the basis of network capacity requirements. Therefore, the artificially created market split of telecommunication providers into specialized services based on active traffic management for the provision of high quality of service levels (e.g. VoIP, IPTV) and passive (TCP-based) best effort Internet is unlikely to remain.

A forward looking economic approach of active traffic management is indicated within a generalized DiffServ architecture based on opportunity costs of traffic capacities and quality of service differentiation. Within this “umbrella” architecture for traffic management, a flexible framework for different traffic quality differentiation strategies is provided. Since its basic characteristic is that all applications are bearing the opportunity costs of their required traffic capacities, the traditional differentiation between managed services and other IP-based Internet services becomes obsolete. In order to guarantee the high quality of VoIP or IPTV, top quality classes can be introduced using the principle of resource reservation with guaranteed end-to-end control. For applications which are less delay sensitive but still require some active traffic management lower traffic quality classes are sufficient, for those applications which are not delay sensitive a “best effort” transmission class may be introduced. In order to provide incentive compatible quality of service differentiation within a generalized DiffServ architecture transmission charges must be monotone increasing with the highest quality class paying the highest transmission charges, and the “best effort” class may be provided for free.
The paper is organized as follows: In the subsequent section 2 the changing role of the IETF standard setting process from designing and implementing the best effort TCP/IP protocol as a universal standard towards a platform for dealing with the increasing need for variety in the design of a QoS differentiated traffic management architecture is demonstrated. The IETF is neither focused on best effort transmissions nor is it conservative, only favoring application innovations at the margins. Instead, in section 3 the IETF’s contributions towards active network management, and subsequently to a more “intelligent” Internet traffic architecture are shown. The IETF’s contributions to a flexible open transmission architecture able to supply the required transmission qualities for the different applications provide the relevant pillars towards a Generalized DiffServ architecture. In section 4 the role of entrepreneurial traffic management within Generalized DiffServ architecture and the division of labor between the IETF and entrepreneurial traffic management is considered. From an economic point of view and in order to enable the development of economically efficient resource allocation mechanisms for traffic capacities it is important to understand the usage of network resources depending on the different transmission qualities and the incentives for investments in transmission network infrastructure. Within the “umbrella” architecture of Generalized DiffServ with the potential to combine basic elements of QoS differentiated traffic architectures a flexible framework for entrepreneurial traffic quality differentiation strategies evolves. Its basic characteristic is market driven network neutrality so that all applications are bearing the opportunity costs of their required traffic capacities. As a consequence an artificial market split between best effort TCP and managed services would conflict with the integrated service approach of the IETF. Within DiffServ and within IntServ and its combinations, best effort transmission service is only one of several classes within the multipurpose Internet architecture. Finally in section 5 the implementation of Generalized DiffServ via Next Generation networks is considered.
2. The changing role of the IETF standard setting process from best effort TCP to Generalized DiffServ Architecture

2.1. The Standard setting process of the IETF

In the past, the Internet Engineering Task Force (IETF) has been the most important standard setting agency for the Internet. The Internet standard setting process described in RFC 2026 (Bradner 1996) differentiates between three stages of maturity: proposed standard, draft standard, and Internet standard, also called standard. Proposed standards are considered to be immature specifications. Since the content of proposed standards may be changed due to practical implementation problems identified or advanced solutions developed, implementation of proposed standards into a disruption-sensitive network is not recommended by the IETF. However, the implementation of a proposed standard is a network provider’s entrepreneurial decision. In contrast, a draft standard indicates a strong belief that the specification is mature and useful. Draft standards are usually considered to be a final specification and the implementation of draft standards into a disruption-sensitive environment is recommended to vendors. An Internet standard (simply referred to as a standard) has a high degree of technical maturity and there is a general belief that the specified protocol provides significant benefit to the Internet community (Bradner 1996, p. 13).

With RFC 6410 of October 2011 the standard setting process has been reformed (Housley et al. 2011). The underlying reason was that in the last decade very few specifications have advanced on the maturity ladder and the vast majority of standards track documents are published as proposed standards. The three stages of maturity have been reduced to two maturity levels. Whereas the requirements for proposed standards have not been changed, the second and third maturity levels have been merged into the Internet standard (Housley et al. 2011). Moreover, specifications not intended to become a standard are considered on the Non-Standards Track Maturity Levels either labeled as “Experimental”, “Informational” or “Historic”. Such specifications may either not be intended to become an Internet standard, or considered not yet ready to enter the standards
track, superseded by a more recent Internet standard, or otherwise fallen into disuse or disfavor (Bradner 1996, p. 13).

The IETF standard setting process is well documented. Subcommittees are called working groups and each proposal (called Internet draft/ID) is published as Request for Comments (RFC). All RFCs have been listed as Internet Standards, Draft Standards, Proposed Standards, Best Current Practice by BCP, Best Current Practice by RFC, Experimental RFCs, or Historic RFCs from the foundation of the IETF in 1986 until the present time.\(^2\) Compared to proposed standards the number of Internet standards is rather small. Moreover, there exists a larger set of non-standards track RFCs.

### 2.2 The best effort TCP/IP standard

The Internet is a challenging battlefield for analyzing the role of standard setting processes in a dynamic environment. Cerf and Kahn (1974) already developed the Transmission Control Protocol/Internet Protocol (TCP/IP) where instead of the transmission network being responsible for reliability the hosts became responsible. The best effort TCP/IP Internet was developed in the narrowband context, when data transmission consisted only of e-mails and small file transfers. There was universal connectivity and a rapid implementation of mandatory compatibility standards by the IETF to exhaust network externalities. Active entrepreneurial traffic management was not relevant. Within the narrowband Internet QoS was not a marketable traffic service concept and no entrepreneurial decisions for active traffic management were required (e-mails, and small file transfers were carried over telephone service networks). The focus was on network externality and universal connectivity. The differentiation between the three stages of the standard setting process was relevant because only phases 2 and 3 provided the final check of acceptance for universal connectivity.

\(^2\) [http://www.rfc-editor.org/rfcxx00.html](http://www.rfc-editor.org/rfcxx00.html)
During the early period of standard setting the focus was on the mandatory introduction of TCP/IP. The traditional best effort TCP/IP assigning all data packages the same priority includes Van Jacobson’s slow start congestion avoidance mechanism described at the February 1987 IETF meeting and published in 1988 (Jacobson 1988). Obligatory requirements for Internet hosts regarding application and support are specified in RFC 1123 (Braden (ed.) 1989a), and regarding communication layers in RFC 1122 (Braden (ed.) 1989b). These RFC documents are issued as official specifications for the Internet community (also called Internet standard or standard) enumerating the standard protocols that a host connected to the Internet must use and incorporating via reference the RFCs and other publications describing the specifications for these protocols. For each protocol an explicit set of requirements is specified, in addition recommendations and possible alternatives are provided. All parties involved in operating hosts fall under these obligations, including vendors, implementers, and users of Internet communications software.

2.3  QoS differentiations as the major driver of the IETF’s changing role

Simcoe (2012) analyzed the IETF’s standard setting procedures, focusing on the impact of the increasing commercialization of the Internet during the period between 1993 and 2003. His basic hypothesis states that due to distributional conflicts and subsequent strategic maneuvering within the IETF the time required to find a consensus would increase and delays in technology adoption could be identified. Simcoe (2012, pp. 315 f.) compared standards track with non-standards track proposals, assuming that non-standards track proposals would create no distributional conflicts, so there would be no correlation between conflict and delay. The development of proprietary technology implemented around the edges and the possibilities of its application within the global TCP/IP based Internet would create incentives for vendors to increase their efforts to participate in the IETF, for example new application services would require new standards for user authentication etc. Detailed committee and proposal-level data were raised for the period between 1993 and 2003 when rapid Internet commercialization significantly changed the size and demographics of the IETF. A large
number of drafts were considered, reflecting the increased demand for new protocols to extend the functionality of the layers near the top of the TCP/IP stack at the edges.3

The basic hypothesis of this paper is that due to the transition from narrowband to broadband Internet the nature of the standard setting process changed dramatically: potentials for variety of QoS differentiation conflict with the implementation of a homogeneous best effort TCP/IP standard, the evolutionary search for a traffic infrastructure which permits alternative QoS implementation strategies. In the meantime, the role of the IETF has changed from developing and enforcing a best effort TCP/IP based traffic infrastructure to a platform for dealing with the increasing need for variety in the design of different traffic management infrastructures and the continuous search for new technological solutions. Moreover, the division of labor between the IETF and entrepreneurial traffic management becomes relevant. Thus cooperation with other standard setting committees (ITU-T etc.) also gains increasing importance.

As early as 1994 the IETF pursued a multipurpose transmission network perspective, cf. RFC 1633 (Braden et al. 1994). As a consequence, technical network neutrality became the wrong reference point, because it is only by means of active traffic management and QoS differentiation of data packet transmission that the different categories of application services can be served within a common Internet infrastructure.

---

3 The TCP/IP protocol stack differentiates between five complementary layers consisting of the physical layer, the data link layer, the Internetwork layer, the transport layer, and the application layer. The top of the TCP/IP stack is in particular the applications layer. The transport layer ensures e.g. reliability of data packet transport, whereas the routing of the packets across the networks is part of the Internetwork layer (Kurose, Ross 2012). For the economic analysis of the allocation of Internet traffic services the disaggregated representation of the Internet as Internet traffic services (Internet access services, Internet backbone services) versus Internet application services (Voice over IP, search machines, content delivery services) seems useful (Knieps, Zenhäusern 2008, p. 122). Internet traffic services use as input not only the transport layer, but also the Internetwork layer and the data link layer (Wolfrum 2013, pp. 145 ff.).
The analysis of this standard setting process during the last two decades shows that neither DiffServ architecture nor IntServ/RSVP has reached the final status of Internet standard. Moreover, important developments of network management design for DiffServ networks did not even reach standards track status, but are considered to be on the non-standards track. This includes the Configuration Guidelines for DiffServ Service Classes. Moreover, Procedures for Modifying the RSVP reached the status of Best Current Practice for the Internet community. The IETF accepts the division of labor between standard setting committees and entrepreneurial traffic management: “However, we expect that network administrators will implement a subset of these classes relevant to their customers and their service offerings: Network administrators may also find it of value to add locally defined service classes, although these will not necessarily enjoy end-to-end properties of the same type”, RFC 4594 (Babiarz et al. 2006, pp. 3 f.). ”Further, much of the details of service construction are covered by legal agreements between different business entities and we avoid this as it is very much outside the scope of the IETF”, RFC 2474 (Nichols et al. 1998, p. 4).

But during the period between 1993 and 2003 the IETF already published important RFCs regarding QoS differentiation.\(^4\) These RFCs created a “silent” potential for basic reforms of Internet traffic architecture and related basic changes of the role of the IETF standard setting process. In the meantime the transition from narrowband to broadband Internet increased the necessity for QoS differentiation, challenging best effort TCP. The search for a QoS differentiated traffic architecture which is open to entrepreneurial variety of QoS differentiation is becoming increasingly important. The IETF is changing towards a platform for the development of standards for basic components for QoS architecture and non-standards track suggestions as thought experiment input for the development of entrepreneurial QoS differentiated traffic architectures.

In this context it is also important to differentiate between standardized basic components for QoS differentiation architecture and the framework design of QoS differentiated architectures based on these standardized components. From the perspective of QoS differentiation the status of a proposed standard seems sufficient for basic components, because their technical feasibility is guaranteed, whereas the universal connectivity check required for the status of an Internet standard becomes superfluous, because implementation is left to the entrepreneurial decisions of traffic network owners. As a consequence the reform of the IETF’s standard setting process can be more rigorous, considering proposed standards as final Internet standards. Moreover, the division of labor between the IETF as the standard setting committee and the entrepreneurial potentials for implementing QoS differentiated traffic architectures and related price and QoS differentiation strategies should be exhausted.

3 Towards a generalized Differentiated Service Architecture and the changing role of the IETF

3.1 QoS differentiation within a generalized DiffServ architecture:
Basic principles

Due to the transition from narrowband access to broadband access, heterogeneous requirements for traffic qualities become increasingly important taking into account the necessities for prioritization of data packets and quality of service guarantees. A transition to active network management and subsequently to a more “intelligent” Internet traffic architecture is required. Although the process of technological innovation regarding the basic elements of QoS differentiated transmission networks was initiated two decades ago, in the meantime the entrepreneurial challenges to implementing QoS differentiated architectures has gained momentum. Therefore, the traditional market split of telecommunications providers into specialized services based on active traffic management for the provision of high quality of service levels (e.g. VoIP, IPTV) and passive (TCP-based) best effort Internet is unlikely to remain stable.
Principles of QoS differentiation within a Generalized DiffServ architecture:

- Lean, non-baroque traffic architecture, avoiding unnecessary complexities as precondition for active entrepreneurial traffic management

- Traffic quality requirements may be derived from application quality requirements, but application service architecture is to be differentiated from active traffic management architecture

- Integrated (multipurpose) service architecture for delay sensitive and non-delay sensitive data packet transmission. Implementation of several traffic classes.

- Sharing of transmission capacity (bandwidth, router) among different traffic classes. Resource reservation for each traffic class versus resource sharing according to the priority principle.

- The principle of technical network neutrality of best effort TCP/IP is replaced by market driven network neutrality. Its basic characteristic is that incentives for discriminating among applications with different traffic requirements are avoided (Knieps 2011, p. 25).

3.2 DiffServ as multipurpose architecture

The role of the IETF changes from enforcing a best effort TCP/IP standard to a platform for the setting of standards for QoS transmission architectures. Although specialized transmission architecture for time sensitive versus time insensitive transmission or “best effort” transmission services seems possible, the basic concept of the IETF during the past decades has been to consider the Internet transmission network as a common transmission network:

“We make another fundamental assumption, that it is desirable to use the Internet as a common infrastructure to support both non-real-time and real-time communication. One could alternatively build an entirely new, parallel infra-
structure for real-time services, leaving the Internet unchanged. We reject this approach, as it would lose the significant advantages of statistical sharing between real-time and non-real-time traffic, and it would be much more complex to build and administer than a common Infrastructure.” RFC 1633 (Braden et al. 1994, pp. 5 f.). The subsequent (proposed) standards as well as the non-standards track (informational) RFCs on QoS differentiated traffic architectures are all based on this basic conviction of the Internet as a multipurpose packet transmission architecture.

The two basic architectures for QoS differentiated packet transmission developed by the IETF consist of (proposed) standards for basic components for IntServ/RSVP and DiffServ architectures that have been developed more than two decades ago. Both architectures are multipurpose architectures differentiating between alternative forms of high quality transmission and lower quality transmission and best effort transmission. The two approaches present two different network concepts.

3.3 Building blocks for traffic classes within DiffServ

DiffServ/DS architecture, focusing on quality of service differentiation of Internet traffic services is described in RFC 2475 (Blake et al. 1998, pp. 9-11). The basic elements of DiffServ architecture are (1) setting bits in an IP header field (DS field) at network boundaries, (2) using these bits to determine how packets are forwarded by the nodes inside the transmission network, and (3) conditioning the marked packets at the network boundaries according to the requirements of each service. Per hop behavior (PHB) determines the specific forwarding treatment for that packet according to the DS Code Point (DSCP) in the IP header.

The network handles packets in different traffic streams by forwarding them using different per-hop-behaviors (PHBs). Many traffic streams can be aggregated to one of a small number of behavior aggregates which are each forwarded using the same PHB, thus simplifying the processing and associated storage, see RFC
According to corresponding proposed standards two different kinds of PHB are specified providing different QoS and subsequently different transmission resource requirements.

(a) Expedited Forwarding / EF PHB

EF is intended to provide a PHB building block for high traffic quality at a given output interface by guaranteeing that the EF aggregate is served at a certain rate (over a suitably defined interval), independent of the offered load of non-EF traffic. Delay and jitter (variation of maximum and minimum delay) are minimized when queuing delays are minimized. The intent of the EF PHB is to guarantee that suitably marked packets usually encounter short or empty queues. If queues remain short relative to the provided buffer space, packet loss is at the same time at a minimum. To ensure short queues it is necessary to ensure that the service rate of EF packets on the output interface exceeds their arrival rate at that interface over long and short time intervals, irrespective of the load of the other (non-EF) traffic. Hence EF is a priority class compared to all other traffic classes.\(^5\) The specification of the standard is considered to be obligatory, although the EF PHB is not intended to become a mandatory part of the Differentiated Services architecture, but depends on the heterogeneous traffic mix of the user side.

(b) Independently assured forwarding /AF PHB

In the proposed standard RFC 2597 (Heinanen et al. 1999) four independent AF classes are defined. Within each AF class an IP packet can be assigned one of three different levels of drop precedence determining the relative importance of the packet within the AF class. A DS node will only reorder IP packets if they do not belong to the same AF class.

In each DS node every class is allocated an amount of forwarding resources (bandwidth, buffer space). The QoS level of forwarding assurance of an IP

\(^5\) See also RFC 3246 (Davie et al. 2002, pp. 2 ff.).
packet depends on the amount of forwarding resources allocated to the respective AF class, the current load of the AF class and (in case of congestion) the drop precedence of the packet.

Application of the standard is not intended to be obligatory, however if an AF PHB group is to be implemented at a DS-compliant node, it must conform to the specifications of RFC 2597. At first glance entrepreneurial flexibility to define traffic class-specific QoS seems to be limited in order to favor the interoperability goal of QoS differentiation among different transmission networks. However, transmission service providers may have incentives to start with a few service classes and improve the degree of differentiation within the entrepreneurial trial and error search process of quality differentiation. Flexibility to increase the number of service classes according to demand side characteristics seems desirable and is also in the spirit of the IETF. DiffServ is not flow based, instead aggregated packet transmission within traffic classes is pursued, which is particularly suitable for large networks. “The defined structure for providing services allows several applications having similar traffic characteristics and performance requirements to be grouped into the same service class. This approach provides a lot of flexibility in providing the appropriate level of service differentiation for current and new, yet unknown applications without introducing significant changes to routers or network configurations when a new traffic type is added to the network” RFC 4594 (Babiarz et al. 2006, p. 11).

3.4 Building blocks for QoS within IntServ/RSVP

To use the Internet as a common infrastructure to support both non-real-time and real-time applications (real-time and all others called best effort) has been pursued in the corresponding RCFs. The parallel infrastructure approach was rejected in an informational RFC 1633 (Braden et al. 1994). The alternative of building an entirely new, parallel infrastructure for real-time service leaving the

---

6 See flexibility regarding the choice of traffic classes in RFC 4594 (Barbiarz 2006, p. 8).
traditional best effort internet unchanged was not pursued. Real-time applications are based on the flow-specific state and resource reservation in the routers. IntServ (proposed) standards support integrated services differentiating between real-time and non-real-time services: best effort service, real-time service, and controlled link sharing require the Resource Reservation Protocol (RSVP) to set up an end-to-end path and to reserve resources along the path before data transmission starts, similar to circuit-switched telephone networks (Braden et al. 1994, Shenker et al. 1997a, Chen, Zhang 2004, pp. 368 f.). RSVP requires admission control and Policy Control before the RSVP process sets up parameters in the packet classifier and packet scheduler to obtain the desired QoS allowing fine granularity. The Integrated Services/RSVP model relies on the exchange of signaling messages between sources and receivers. The basic principle of the Internet integrated services framework is to provide the ability for applications to choose among QoS controlled delivery services (such as Controlled-Load and Guaranteed QoS) for their data packets. There are two service definitions as proposed standards within IntServ/RSVP:

(a) Guaranteed services

Guaranteed services provide a firm guaranteed delay bound, intended for applications that have hard real-time requirements in proposed standard RFC 2212 (Shenker et al. 1997b, p. 3). Guaranteed transmission service based on dedicated bandwidth for specific traffic flows produces a delay-bounded service where queuing losses within these flows can be effectively avoided.

(b) Controlled-Load Service

Controlled-Load Service is able to support a broad class of applications, in particular “adaptive real-time applications” which work sufficiently well on unloaded networks but degrade quickly under overloaded conditions, in proposed standard RFC 2211 (Wroclawski 1997, p. 2).

IntServ requires, firstly that individual network elements (subnets and IP routers) along the path followed by an application’s data packets must support
mechanisms which control the quality of service of delivery for these packets; secondly, a way to communicate the requirements of applications to network elements along the path and the provision of QoS management information between network elements and the applications. This can be provided by a resource reservation setup protocol such as RSVP.

A logical separation exists between QoS control services and RSVP setup mechanisms. In addition to the data used to directly guarantee QoS control services, RSVP also carries authentication, accounting and further information needed to manage the admission control for use of QoS services (Wroclawski 1997, p. 1 f.). IntServ/RSVP seems to have particular advantages regarding time-sensitive applications due to its flow-dependency, combined, however, with the disadvantage of high signaling costs and subsequent relative low scalability.

RSVP is a signaling protocol which signals per-flow resource requirements to network elements within the IntServ domain expressing service types (guaranteed services, controlled-load and best effort transmission), quantifying resource requirements and determining the availability of the requested resource at the network elements (admission control) in proposed standard RFC 2996 (Bernet et al. 2000, pp. 2 ff.). Subsequently further developments and modifications in several proposed standards have been published. Their focus is e.g. on the reduction of bandwidth of a reservation flow in proposed standard RFC 4495 (Polk et al. 2006) or the implementation of RSVP signaling, if the receiver is not RSVP-capable by RSVP Receiver Proxy in proposed standard RFC 5946 (Le Faucheur et al. 2010). The question arises whether and to what extent this resource intensive signaling protocol is required for guaranteeing real-time traffic and to what extent non-flow based packet transmission within a premium class of DiffsServ seems sufficient.

7 “RSVP is not itself a routing protocol; RSVP is designed to operate with current and future unicast and multicast routing protocols” RFC 2205 (Braden et al. 1997, p. 4).
3.5 Integration of DiffServ and IntServ within the Generalized DiffServ architecture

DiffServ and IntServ architectures can be implemented in combination:

“Our goal is to enable seamless inter-operation. As a result, the network administrator is free to choose which regions of the network act as Diffserv regions. In one extreme the Diffserv region is pushed all the way to the periphery, with hosts alone having full Intserv capability. In the other extreme, Intserv is pushed all the way to the core, with no Diffserv region.” RFC 2998 (Bernet et al. 2000, p. 5).

For aggregation of individual reserved sessions into a DiffServ class see RFC 3175 (Baker et al. 2001). In small and bandwidth-constrained networks for limited number of flows RSVP may be used, however, concern arises over the scalability on large networks where therefore aggregation of reservations seems necessary, cf. RFC 3175 (Baker et al. 2001). Moreover for VoIP signaling protocols see RFC 4594 (Babiarz 2006, p. 11). The goal is the aggregation of individual reserved sessions into a common class to achieve scalability.

The problem with many small reservations is that each reservation requires a non-trivial amount of message exchange, computation, and memory resources in each router along the path. RSVP Version 1 did not find a way to aggregate sessions because there was no clear way to classify the aggregate. Within the DiffServ architecture aggregate reservations from ingress to egress routers of the DiffServ network or aggregation region (“aggregating” router, and “deaggregating” router) can be marked with a given DSCP (e.g. EF and classified as a premium traffic class) where each aggregate reservation carries similarly marked packets from a large number of flows. One or more DiffServ DSCPs are applied to identify traffic with the same aggregate reservations and one or more DiffServ PHBs are used to provide the required forwarding treatment to this traffic class. Within an aggregation region the interior (core) routers only transmit data packets in the DiffServ style without flow based reference (end-to-end RSVP messages are hidden from the interior routers). One of the major ben-
Benefits of aggregation is the reduction of message processing cost in the aggregation region. The most simple policy would be to map all end-to-end reservations onto a single aggregate reservation, a single DSCP; another possibility would be to map guaranteed service end-to-end reservations onto one DSCP in the aggregation region and controlled-load end-to-end reservations onto another DSCP. How much bandwidth should be allocated to an aggregated reservation at any given time is considered to be proprietary to the service provider’s business policy.\(^8\)

The following may be a scalable and efficient solution: In access networks IntServ may be implemented, then it may be aggregated into one or more DiffServ classes within the core network. By merging Integrated Services/RSVP into DiffServ architecture DiffServ mechanisms can be applied to aggregate Integrated Services/RSVP in the core of the network as described in RFC 2475 (Blake 1998, p. 11). End-to-end QoS guarantees can be provided by introducing a sender-initiated resource reservation mechanism over DiffServ networks with either absolute QoS guarantees or relative services guarantees based on statistical values (Zhang, Mouftah 2001).

Summing up, several alternative architectures for data packet transmission have been developed within the IETF:

(1) one extreme is best effort congestion control, which is exclusively performed by the communicating edges and results in average traffic quality,

(2) the other extreme is represented by the implementation of IntServ/RSVP, admission control via resource reservation along a transmission path, ensuring guaranteed QoS (traffic quality)

(3) via means of prioritization, DiffServ enables different QoS for different traffic classes, consisting of traffic aggregates; premium classes satisfy the demand for delay sensitive applications. Traffic quality can be ensured based on statistical probabilities

\(^8\) See Aggregation of RSVP for IPv4 and IPv6 Reservations, in Proposed Standard RFC 3175 (Baker et al. 2001).
(4) the aggregation of IntServ/RSVP-based traffic flows into aggregates of one or more DiffServ traffic classes.

3.6 The evolutionary perspective of the IETF regarding QoS class differentiation

The type of service (TOS) field of an IP packet providing the possibility of a choice between minimized delay, maximized throughput, maximized reliability, minimized monetary cost and normal service has already been specified in a proposed standard in RFC 1349 (Almquist 1992). Guidelines for DiffServ service classes have been developed in the informational RFC 4594 (Babiarz et al. 2006). The focus is on traffic requirements as precondition for providing different services: “A ‘service class’ represents a set of traffic that requires specific delay, loss, and jitter characteristics from the network.” RFC 4594 (Babiarz et al. 2006, p. 5). Different traffic groups are considered. The network control traffic group takes into account network control for routing and network control function. OAM (Operations, Administration and Management) is focused on network configuration and management functions.

“The user/subscriber traffic group is broken down into ten services classes to provide service differentiation for all the different types of applications/services” RFC 4564 (Babiarz et al. 2006, p. 12). Service classes provide indication of the traffic forwarding treatment to meet user or application (or network) requirements. The different service classes are considered as reference names with QoS behaviors that are optimized for the particular application types they support not implying any priority ordering (Babiarz et al. 2006, p. 18).” However, the underlying resource allocation problem permits a comparison of the different opportunity costs of transmission services in different traffic classes, implicitly revealing some priority ordering of different quality classes (see section 4.3).

---

9 Service class names and application examples are listed in Figure 3 in RFC 4594 (Babiarz et al. 2006, p. 19).
The standard service class contains traffic which has not been classified into one of the other service classes in the DiffServ network domain. It has minimum bandwidth guarantee and at least a small percentage of forwarding resources are a guaranteed minimum. The standard service class provides the DiffServ network domain “best-effort” forwarding behavior, cf. RFC 4594 (Babiarz et al. p. 47). The standard service within DiffServ architecture is still part of active traffic management, compared to passive best effort TCP/IP traffic management.

Nevertheless, the IETF considers the traffic classes to be open to enable a wide variety of application services and network configurations. Network administrators may only provide a subset of the defined service classes, but may also introduce other traffic service classes in their networks, see RFC 4594 (Babiarz et al. 2006, p. 16).

These user service classes can be grouped into a smaller number of application categories. Due to the different traffic characteristics of the applications, control functions and the required flow behavior for some application categories more than one service class was required to provide service differentiation within that category. The allocation of network resources (bandwidth, router capacity) depends on the required traffic services needed for the different application services.

“Service class definitions are based on the different traffic characteristics and required performance of the applications/services. This approach allows us to map current and future applications/services of similar traffic characteristics and performance requirements into the same service class” see RFC 4594 (Babiarz et al. 2006, p. 3).

It can be concluded that the IETF is neither focused on best effort transmission, nor conservative, only favoring application innovations at the margins. Instead, the IETF is designing flexible open transmission architecture able to provide the required transmission qualities for the different applications: this can be termed generalized DiffServ architecture.
4. Entrepreneurial traffic management within Generalized DiffServ Architecture

4.1. Division of labor between the IETF and entrepreneurial traffic management

The entrepreneurial role of network resource allocation has already been pointed out by the IETF:

“The exact policies used in determining how much bandwidth should be allocated to an aggregate reservation at any given time are beyond the scope of this document, and may be proprietary to the service provider in question.” RFC 3175 (Baker et al. 2001, p. 20).

“This document provides guidelines for network administrators in configuring their network for the level of service differentiation that is appropriate in their network to meet their QoS needs. It is expected that network operators will configure and provide in their networks a subset of the defined service classes. Our intent is to provide guidelines for configuration Differentiated Services for a wide variety of applications, services, and network configurations. In addition, network administrators may choose to define and deploy other service classes in their network”. RFC 4594 (Babiarz et al. 2006, p. 16).

“The authors hope that these Diffserv “project plans” will provide a useful guide to Network Administrators in the use of Diffserv techniques to implement quality-of-service measures appropriate for their network’s traffic”, RFC 4594 (Babiarz et al. 2006).

From an economic point of view and in order to enable the development of economically efficient resource allocation mechanisms for traffic capacities it is important to understand the usage of network resources depending on the different transmission qualities and the incentives for investments in transmission network infrastructure.
4.2 The economics of active traffic management within Internet traffic service networks

Active entrepreneurial traffic management is based on the opportunity costs of capacity usage. The transmission services provided, the choice of traffic classes and their quality characteristics depend endogenously on the requirements of the application providers and users. A traffic service network provider may either sell transmission services only or provide a bundled “end-to-end service” (including application services). Under competition, due to arbitrage conditions the traffic service provider is indifferent between providing traffic service only or “end-to-end service”.

The opportunity costs of capacity usage depend on the specific implementation of traffic architecture. Opportunity costs are different within specialized best effort transmission networks, specialized flow based (circuit switched equivalent) networks and within multipurpose, multi-class transmission networks. Within a Generalized DiffServ architecture opportunity costs arise either due to rivalry over (short-run) available capacity among different traffic classes or due to sharing of capacity among different classes according to priority rules according to interclass externalities pricing.

Price and quality differentiation strategies should not be regulated in such a way that entrepreneurial incentives for market driven network neutrality cannot evolve. If any application is charged according to the opportunity costs of required traffic capacities, incentives for Internet traffic service providers to discriminate between various network applications disappear (Knieps 2011, p. 25). Within the Generalized DiffServ architecture the principle of technical network neutrality originating from best effort TCP/IP architecture is replaced by the principle of market driven network neutrality.
4.3 QoS differentiation and traffic class pricing based on the opportunity costs of capacity usage in the Generalized DiffServ model

An incentive compatible QoS differentiated pricing model is required for the implementation of active traffic management in a generalized DiffServ architecture. Within the “umbrella” architecture of Generalized DiffServ, a flexible framework for different traffic quality differentiation strategies evolves. Its basic characteristic is market driven network neutrality such that all applications are bearing the opportunity costs of their required traffic capacities. In order to guarantee the high quality of VoIP or IPTV, top quality classes can be introduced using the principle of (aggregated) resource reservation with guaranteed end-to-end control within premium DiffServ classes. For applications which are less delay sensitive but still require some active traffic management lower traffic quality classes are sufficient. For those applications which are not delay sensitive a “best effort” transmission class may be introduced. In order to provide incentive compatible quality of service differentiation within a generalized DiffServ architecture transmission charges must be monotone increasing with the highest quality class paying the highest transmission charges, and the “best effort” class may be provided for free. Whereas in the context of pure DiffServ architecture such a pricing scheme has been developed based on interclass externality pricing (Knieps 2011), within a Generalized DiffServ architecture the more general principle of rivalry for network resources used for different traffic classes can be applied.

4.4 Instability of artificial market split between best effort TCP and managed services within Generalized DiffServ Architecture

An artificial market split into managed services and best effort TCP has not been proposed by the IETF, on the contrary, this would even conflict with the IETF’s integrated service approach. In the context of DiffServ and IntServ and its combinations, best effort transmission service is only one of several classes within the multipurpose Internet architecture.
The use of best effort TCP for different traffic qualities on the one hand and exogenously chosen managed service classes on the other hand artificially restricts the entrepreneurial flexibility to consistently apply the principle of all applications bearing the opportunity costs of their capacity usage. In particular, the entrepreneurial freedom to meet the demand for further traffic classes is artificially limited. The consequence of Generalized DiffServ architecture is that economic incentives for market driven network neutrality arise. Thus opportunity cost based transmission pricing, depending on required transmission capacities and interclass externalities, is developed. An artificial market split between best effort TCP and quality guaranteed managed services becomes unstable as a consequence of market driven network neutrality.

4.5 Interoperability between generalized DiffServ networks

One extreme solution has been the proposal to integrate all or several transmission network owners into one centralized network with one standardized Internet traffic architecture and integrated QoS differentiated capacity allocation decision (Li et al. 2004, p. 93). However, a centralized integrated solution is contrary to the transmission service providers’ competitive entrepreneurial freedom to make their own business decisions.

A market based superior solution pursued by the IETF is to leave the entrepreneurial choice how to use the proposed standard components and the subsequent network architecture decision to the individual transmission network provider. This choice depends in particular on the business models of the different transmission service network providers. A large variety of different QoS differentiation strategies may evolve: specialized networks focusing on different applications, e.g. high quality transmission based interactive videogames for specific user groups, or high quality based universal connectivity Voice over IP, or medium quality transmission based data distribution services, or multipurpose networks with delay sensitive and delay insensitive transmission services.
The advantage is the entrepreneurial flexibility to develop one’s own QoS transmission architecture (based on required proposed standards for components of QoS architecture) and subsequent innovations for network evolution. The goal is to change the reference point of mandatory standardization from a best effort TCP based transmission network to an entrepreneurial choice within the Generalized DiffServ Architecture Framework with reference to QoS differentiated data packet transmission. This framework should be general enough to allow entrepreneurial flexibility in the implementation of the Generalized DiffServ Architecture Framework (which elements remain passive, which elements become active). As a consequence the evolutionary search for improving QoS differentiated compatibility between transmission network operators is not hampered but stimulated.

5. Implementation of Generalized DiffServ via Next Generation Networks

The focus of Generalized DiffServ networks is on the transmission of data packets for heterogeneous application services, such as telecommunications services (e.g. Voice over IP), media services (e.g. IPTV), and data transmission services (e.g. e-mail, document transfer). Although the design of traffic classes is taking into account the transmission quality requirements of the different application hardware and software beyond the scope of DiffServ architecture, the IETF has nevertheless also worked on standards track RFCs and informational non-standards track RFCs dealing with application protocols.

Due to the increasing role of the provision of telecommunications, media and data transmission via an integrated IP based communications infrastructure it is not surprising that the International Telecommunication Union (ITU) and its permanent committee ITU-T, the ITU Telecommunication Standardization Sector, are providing recommendations regarding traffic control and congestion
control in IP based networks.\textsuperscript{10} According to ITU-T (2004, p. 2) Next Generation Networks are characterized as follows: “A packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It enables unfettered access for users to networks and to competing service providers and/or services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users.”

Basic characteristics of Next Generation Networks/NGN are (1) that all transmission is IP based, (2) the decoupling of application services and transport in such a way that service related functions are independent from transport technologies, (3) broadband capabilities with end-to-end QoS differentiation and different treatment of data traffic within its backbones, (4) migration of a large number (open set) of services, including voice services to be served by the NGN transport infrastructure (migration of voice from a circuit switched architecture (PSTN) to VoIP) removes the voice switching infrastructure from the exchange, and (5) the support of both existing and “NGN aware” end terminal devices.

The basic principles of NGN are in the spirit of generalized DiffServ architecture. In particular an integrated multipurpose traffic service network on which an open set of application services can be provided by competitive application service providers. QoS differentiation in the traffic service network and QoS differentiation in the application services can be implemented by different actors. However, adequate interfaces are required to achieve the desired end-to-end quality of service.

\textsuperscript{10} The status of ITU-T recommendations is characterized by the ITU-T as follows: “Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure, e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met.”(ITU-T 2004, p. ii).
5.1 QoS differentiation within NGN transmission networks

Recommendations concerning QoS differentiation in transmission networks are strongly based on the InTServ/RSVP and DiffServ architecture developed by IETF (ITU-T 2002, p. 1). In the meantime the Multiprotocol Label Switching Architecture (MPLS) in Proposed Standard RFC 3031 (Rosen et al. 2001) and its combination with Differentiated Services in RFC 3270 (Le Faucheur (ed.) et al. 2002) is also being considered as an innovative traffic routing mechanism within NGN (Sietmann 2009, p. 3). MPLS is path oriented, in such a way that all packets belonging to a particular “Forwarding Equivalence Class” (FEC) which travel from a given node will follow the same path. In contrast to connectionless IP transmission with MPLS an absolute QoS guarantee can be provided (Le Faucheur (ed.) et al. 2002, p. 4). Moreover, congestion management due to an active time dependent choice of path can be improved. The entrepreneurial question arises whether and to what extent this resource intensive signaling protocol is required for guaranteeing real-time traffic and to what extent IP based packet transmission within a premium class of DiffServ seems sufficient.

5.2 Complementary application services

Complementary application services are to a large extent standardized by the IETF. For example, the Session Initiation Protocol (SIP) runs on top of several different transport protocols. “SIP invitations used to create sessions carry session descriptions that allow participants to agree on a set of compatible media types”, RFC 3261 (Rosenberg et al. 2002, p. 1).

According to the Proposed Standard RFC 3588 Diameter Base Protocol (Calhoun et al. 2003, p. 1) the diameter base protocol provides an Authentication, Authorization and Accounting (AAA) framework pursued for applications such as network access or IP mobility.

On competitive transmission service networks, bundling between transmission services and application services should be allowed. Although Next Generation
Networks (NGN) may be implemented according to a Generalized DiffServ architecture principle (lean NGN), over-complex and baroque Internet traffic architecture should be avoided.

References

Almquist, P. (1992), Type of Service in the Internet Protocol Suite, RFC 1349

Babiarz, J. et al. (2006), Configuration Guidelines for DiffServ Service Classes, RFC 4594

Baker et al. (2001), Aggregation of RSVP for IPv4 and IPv6 Reservations, RFC 3175

Bernet et al. (2000), A Framework for Integrated Services Operation over Diffserv Networks, RFC 2998

Bernet et al. (2002), An Informational Management Model for Diffserv Routers, RFC 3290

Blake, S. et al. (1998), An Architecture for Differentiated Services, RFC 2475

Braden, R. (Ed.), (1989a), Requirements for Internet Hosts – Application and Support, RFC 1123

Braden, R. (Ed.), (1989b), Requirements for Internet Hosts – Communications Layers, RFC 1122

Braden et al. (1994), Integrated Services in the Internet Architecture: an Overview, RFC 1633


Calhoun, P. et al. (2003), Diameter Base Protocol, RFC 3588


Davie, B. et al. (2002), An Expedited Forwarding PHB (Per-Hop Behavior), RFC 3246

Heinanen, J. et al. (1999), Assured Forwarding PHB Group, RFC 2597

Housley, R. et al. (2011), Reducing the Standards Track to Two Maturity Levels, RFC 6410


Knieps, G., Zenhäusern, P. (2008), The fallacies of network neutrality regulation, Competition and Regulation in Network Industries, 9/2, 119-134


Le Faucheur, F. et al. (2010), Resource Reservation Protocol (RSVP) Extensions for Path-Triggered RSVP Receiver Proxy, RFC 5946

Li, T., Iraqi, Y., Boutaba, R. (2004), Pricing and admission control for QoS-enabled Internet, Computer Networks, 46/1, 87-110

Nichols, K. (1998), Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers, RFC 2474

Rosen et al. (2001), Multiprotocol Label Switching Architecture, RFC 3031

Rosenberg et al. (2002), SIP: Session Initiation Protocol, RFC 3261

Shenker et al. (1997a), General Characterization Parameters for Integrated Service Network Elements, RFC 2215

Shenker et al. (1997b), Specification of Guaranteed Quality of Service, RFC 2212


Als Diskussionsbeiträge des
Instituts für Verkehrswissenschaft und Regionalpolitik
Albert-Ludwigs-Universität Freiburg i. Br.
sind zuletzt erschienen:


104. **G. Knieps**: The Different Role of Mandatory Access in German Regulation of Railroads and Telecommunications, erschienen in: Journal of Competition Law and Economics, Vol. 2/1, 2006, S. 149-158


108. **H.-J. Weiß**: Die Potenziale des Deprival Value-Konzepts zur entscheidungsorientierten Bewertung von Kapital in liberalisierten Netzindustrien, Juni 2005


121. **G. Knieps**: Verkehrinfrastrukturen zwischen Wettbewerb und Regulierung, erschienen in: Wirtschaftspolitische Blätter, 56/1, 2009, S. 11-21


124. **G. Knieps**: Qualitäts- und Preisdifferenzierung im Internet, erschienen in: Jörn Kruse, Ralf Dewenter, Hrsg., Wettbewerbsprobleme im Internet, Nomos Verlag, Baden-Baden, 2009, S. 103-113


141. **G. Knieps, P. Zenhäusern**: The reform process of the railway sector in Europe: A disaggregated regulatory approach, December 2011


143. **G. Knieps**: Europäische Telekommunikationsregulierung: Quo vadis?, Januar 2013


145. **G. Knieps**: Der Irrweg regulatorischer Marktpartizipation: Zur Novelle des Personenbeförderungsgesetzes in Deutschland, erschienen in: Zeitschrift für Verkehrswissenschaft, 84/1, 2013, S. 69-77

146. **G. Knieps**: Market versus state in building the aviation value chain, Mai 2013