Internet services beyond best effort

The packet-based data transmission of the Internet allows the multiplexing of a variety of simultaneous connections originating from sources with different characteristics (e.g. voice, video, data). Traditional Internet is based on a best effort service, whereas future Internet services make more and more special demands on the communication network. This paper presents a general overview of fundamental mechanisms for guaranteeing Quality of Service (QoS) in the current and future Internet. After introducing general QoS concepts, basic mechanisms in network routers are investigated. Furthermore, the two basic architectures of Integrated Services (IntServ) and Differentiated Services (DiffServ) are discussed in detail. Finally, the analytical investigation of a special Voice-over-IP scenario demonstrates the applicability of the relevant QoS concepts and their positive consequences with respect to the quality of this voice service. A short overview of current hot research topics concludes the paper.

Keywords: quality of service; Internet; TCP/IP; integrated services; differentiated services; voice over IP; resource reservation; admission control

Internet-Dienste jenseits von Best Effort.

Die paketbasierte Datenübertragung über das Internet ermöglicht das Multiplexen einer Vielzahl gleichzeitiger Verbindungen aus unterschiedlich charakterisierten Datenquellen (z. B. Sprache, Video, Datentransfer). Während im traditionellen Internet allerdings die Übertragung auf dem Best Effort-Prinzip basiert, stellen die zukünftig angebotenen neuen Internet-Dienste laufend mehr spezielle Anforderungen an das Kommunikationsnetz. Dieser Artikel gibt einen allgemeinen Überblick über fundamentale Mechanismen zur Garantie einer gewünschten Übertragungsgüte (Quality of Service, QoS) im heutigen und zukünftigen Internet. Nach der Einführung allgemeiner QoS-Konzepte liegt der Fokus zunächst auf den Grundmechanismen innerhalb der Netzrouter. Sodann werden die beiden grundlegenden Architekturen der Integrated Services (IntServ) und Differentiated Services (DiffServ) im Detail behandelt. Schließlich demonstriert die genauere Untersuchung eines speziellen Voice-over-IP-Szenarios die Anwendbarkeit der eingeführten Prinzipien und ihre positiven Auswirkungen auf die Qualität des betrachteten Sprachdienstes. Ein kurzer Überblick über aktuell behandelte Forschungsthemen rundet den Artikel ab.

Schlüsselwörter: Quality of Service; Internet; TCP/IP; Integrated Services; Differentiated Services; Voice over IP; Ressourcenreservierung; Zugangskontrolle

E. PLASSER, T. ZIEGLER, P. REICHL, G. MITTENECKER OVE

1. Introduction

The basic need for Quality of Service (QoS) comes from the fact that any arrangement of providing a fixed one-size-fit-all service to all customers is insufficient. Users employ different applications, for example Voice over IP (VoIP), Virtual Private Networks (VPNs), Web-browsing, and mission-critical e-commerce, causing the need for a variety of services to be provided by the network. Different services require different treatment of data packets in the network. Hence, QoS is a term referring to technologies that classify network traffic and ensure that some of that traffic receives preferential handling. Customers will be enabled to subscribe different services in terms of throughput, loss, response time, delay and/or delay jitter (the so-called QoS parameters) and will be charged for the actual level of service provided. Literature about QoS can be found in (*Ferguson, Huston, 1998; Siegel, 2000; Greenville Armitage, 2000; Uless Black, 2000; Kilkki, 1999*).

The implementation of the QoS concept began with the emergence of X.25 and its improvements in the late 1960s and early 1970s. The goal was to create a common platform on which all users with a special interface could participate with their data traffic. The design of X.25 allowed users to request certain levels of service, for the first time.

The next steps in providing QoS were the introduction of the Frame Relay (FR) and Asynchronous Transfer Mode (ATM) technologies. FR gained much popularity as a method for interconnecting widely separated geographic locations, but has limited QoS capabilities. ATM was planned to transport as different traffic types as voice, data, video linked with QoS guarantees in a single generic technology. Because ATM was expected to solve many problems in delivering different services, it has seen extensive specifications, standards and research. ATM, however, suffers from the fact that specifications were only partially implemented.

Originally, the Internet Protocol (IP) offered only one service – "best effort". Best effort means that each user is provided service and gets a fair share of the available network resources. On the other hand, no promises can be made concerning QoS guarantees like an upper bound on end-to-end delay, minimum throughput and delay jitter.

Internet traffic is solely controlled by congestion control mechanisms at hosts. Congestion control adjusts the sender's transmission rate into the net as a function of the network load.

PLASSER, Erich, Dipl.-Ing., DWT., ZIEGLER, Thomas, Dipl.-Ing. Dr., REICHL, Peter, Dipl.-Math. Dr., MITTENECKER, Georg, Dipl.-Ing., ftw. Forschungszentrum Telekommunikation Wien, Techgate, Donau-City-Straße 1/3, A-1220 Wien.

If the network load is high, and the available capacity diminishes, the sending rate is reduced. If the network load is low, the sending rate is increased correspondingly. Congestion control is both a means to avoid congestion collapse and to guarantee high network utilization. For the Internet, the Transmission Control Protocol (TCP) congestion control (*Jacobson, 1988; Stevens, 1994*) is of central importance. TCP congestion control uses packet loss as an indication to congestion and adjusts a window of packets as a function of the number of lost packets during the last round trip time. The effective TCP sending rate can be estimated as window size divided by round trip time.

Such an architecture has significant technical advantages in terms of simplicity and high link utilization, is perfectly suited for classical Internet applications like e-mail, FTP (File Transfer Protocol) and Telnet, but lacks the facilities to provide service guarantees. Although a rudimentary QoS support (Type of Service bits in the IP header) was intended from the very beginning, this was hardly used in actual implementations. Nowadays users are no longer satisfied with simple applications like FTP or Telnet alone. Some of the new Internet applications require significant bandwidth. Others, like IP telephony and other real-time applications have strict constraints on timing requirements. These applications require network services beyond the simple "best effort" service that IP delivers.

2. Basic QoS mechanisms in IP

The goal in a QoS-enabled environment is to allow for predictable service delivery to certain classes or types of traffic regardless of other traffic flowing through the network at any given time. We describe the accomplishment of a QoS-connection to achieve this goal in a very generic way by emphasizing on two fundamental steps.

(1) Admission control and resource allocation:

Before the user may send data packets the network has to perform admission control. Given a certain load situation and QoS demand of a flow, admission control determines whether the flow's demand can be granted without violating the service guarantees of existing flows. The decision about admission can be made anywhere in the network. Nevertheless, the enforcement about admission control is delegated to a policer at an edge router (Fig. 1). If the flow is admitted it shares the available network resources with many other flows. Thus it is important to allocate to each flow or each group of flows resources according to predefined QoS guidelines. In most cases resource allocation is made for bandwidth and buffer space, enabling the network service provider to deliver QoS guarantees.

(2) Maintaining the arranged network resources during the lifetime of a flow:

During the packet flow it is to assure that the adjusted conditions are kept and the flow's parameters are maintained by

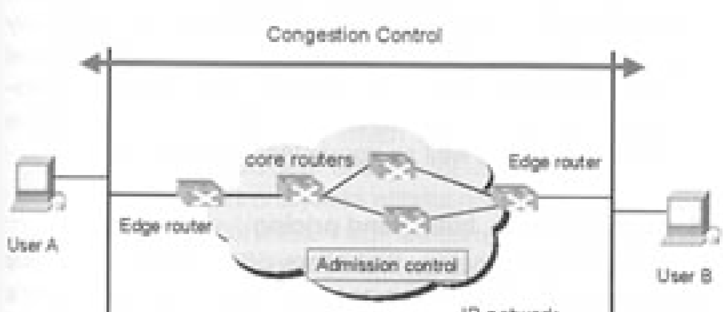

Fig. 1. QoS-connection set up

the router components: classifying, policing, queue management, scheduling.

The following section explains this second step in more detail.

3. Router components in QoS-enabled networks

The routers are the elements in a network where resources (bandwidth, buffer space) are shared by the different traffic types and where algorithms can be adjusted according to the expected different handling. A traditional router mainly focuses on where to send packets. It makes forwarding decisions based on the destination address in each packet and locally held forwarding information tables. QoS-enabled routers must additionally provide control of when to send packets, which packets to send and which packets to drop in case of congestion. Figure 2 shows the components in a router necessary for QoS handling which are part of a router's forwarding process. The following subsections address these and several other associated router components.

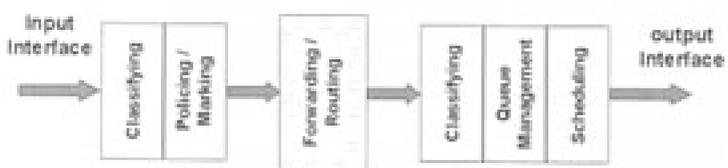

Fig. 2. Forwarding process within a router

3.1 Classification

Receiving a packet, a classifier determines which traffic class the packet belongs to based on the content of one or several fields in the packet header. If several fields are needed, this process is called Multifield Classification (MF). Classifiers are used to pass packets matching some specified rule to a traffic conditioner (which is a term for metering, shaping, policing) (*Blake et al., 1998*) as well as to pass packets to a specified queue management mechanism. Figure 3 shows how a classifier assigns different packets to different queues.

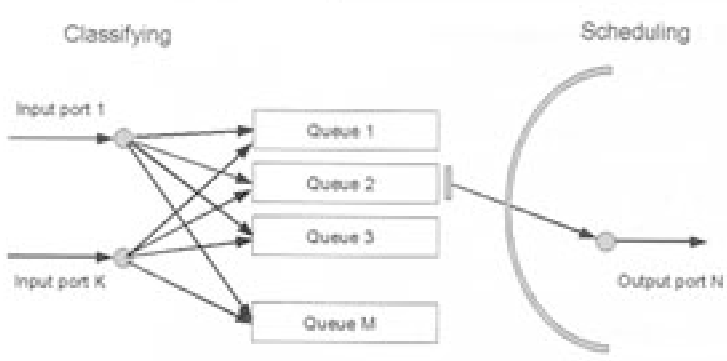

Fig. 3. Router output port architecture

3.2 Policing and marking

In most cases policing is about the supervision of whether the incoming traffic is compliant with pre-negotiated policies (then it is called traffic policing). In general the term policing can be applied to a broad range of rule systems. The term also appears at shaping, dropping, authentication, billing and may involve two (bilateral) or more (multilateral) parties.

Every traffic class has certain limits to its allowable temporal behavior – for instance, a rate limit specified by the number of ticular hop by the service time of the non-voice packet, i.e. by $\tau = 8\,L/B = 1.2$ ms. In the worst case, this happens at each of the n hops on the diameter of the network, yielding a total jitter J experienced by our voice packet of $J = n \cdot \tau$. For diameter $n = 3$, this e.g. corresponds to $J = 3.6$ ms, whereas for $n = 10$ we get $J = 12$ ms.

How many parallel voice connections can at least be guaranteed under these assumptions? Assume, packets from N parallel connections are queueing simultaneously at the first hop. The total jitter under VW (interference plus queueing delay at the first hop) is bounded to be $\Delta = 20$ ms, therefore any voice packet at the first hop should experience a queueing delay caused by other voice packets of no more than $\delta = \Delta - J$. Therefore, the maximal number N of parallel VoIP connections is bounded by $N \cdot 8\,s/B \le \delta$, yielding $N \le \delta \cdot B/8\,s$.

Summarizing this calculation, the maximal number of possible VoIP calls under VW equals

$$N = \frac{\delta \cdot B}{8s} = \frac{(\Delta - J) \cdot B}{8s} = \frac{(\Delta - (n \cdot \tau)) \cdot B}{8s} = \frac{(\Delta - (n \cdot 8 \cdot L/B)) \cdot B}{8s}$$

$$= \frac{\Delta \cdot B - (n \cdot 8 \cdot L)}{8s} = \frac{\Delta \cdot B}{8s} - \frac{n \cdot L}{s} = N_{\max} - \frac{L}{s} \cdot n$$

with Ethernet bandwidth $B = 10$ Mbit/s, maximal length of an Ethernet packet $L = 1\,500$ bytes, network diameter n, G.711 packet size $s = 200$ byte and codec packet delivery time $\Delta = 20$ ms. Therefore, numerically a QoS-enabled (VW) network of diameter n can accommodate up to $125 - 7.5 \cdot n$ calls, without posing any restrictions to the additional (background) data traffic. For a typical network diameter of five hops (*Jacobson, Nichols, Poduri, 2000*), this yields e.g. a number of 87 phone calls.

6. Outlook on QoS related research topics

Ongoing research in the area of IP performance evaluation at ftw (Forschungszentrum Telekommunikation Wien) deals with the enabling technologies to make QoS work in the next generation Internet. This includes the modeling of Internet traffic characteristics, analysis and design of QoS mechanisms as well as the question how to quantify the QoS guarantees provided by DiffServ given a certain network topology, load situation and traffic characteristic. Performance evaluation is based on using theoretical approaches like queuing theory and control theory as well as extensive simulations and measurements.

Research on traffic control evaluates the performance of end-to-end congestion control, queue management, scheduling, traffic conditioning, optimizes parameter settings and proposes new mechanisms in case existing solutions exhibit sub-optimal performance. As an example the work on optimizing the setting of RED parameters (see Sect. 2.4) to avoid poor link utilization and service discrimination between in-profile and out-of-profile packets due to an oscillating queue can be mentioned (*Ziegler, Brandauer, Fdida, 2001*).

Another important area of research at the ftw addresses Internet charging, with special emphasis on pricing, billing and accounting concepts. This includes theoretical and practical investigation and analysis of existing approaches and their improvement as well as the development of new ideas on a technically as well as economically sound basis. Auction-type schemes in multi-provider environments, the Cumulus Pricing Scheme (*Reichl, Stiller, 2001*), pricing and time-scales, the edge pricing paradigm and the emerging importance of pricing and QoS in mobile IP networks are topics within this area.

7. Conclusions

This paper presents a concise overview of fundamental concepts for providing Quality of Service in the Internet of today and tomorrow. After introducing basic QoS-related aspects of the IP protocol, we focus on relevant mechanisms within the routers, ranging from classification, policing, marking and traffic shaping strategies over queue management and scheduling to authentication, billing and pricing. At the moment, basically two QoS architectures have been established: the Integrated Services (IntServ) proposal uses advanced reservation of the required resources and therefore is able to provide strict QoS guarantees, whereas the Differentiated Services (DiffServ) architecture with its focus on traffic aggregation and the definition of Per-Hop-Behaviours is much better in terms of scalability, but less flexible in satisfying individual QoS requirements. For both of these central approaches, the most important service classes are introduced in some detail, before we present an example for the practical application of these concepts in the case of VoIP. Finally, a brief overview of current QoS-research at the ftw. provides insights into the related fields that will be investigated further.

Abbreviations

A0	Project within FTW, including QoS research
AF	Assured Forwarding
AQM	Active Queue Management
BiPA	Billing, Pricing and Accounting in the Internet
CL	Controlled Load
DiffServ	Differentiated Services
DS	DiffServ Field
DSCP	DiffServ Code Point
EF	Expedited Forwarding
FIFO	First In First Out
FR	Frame Relay
FTP	File Transfer Protocol
ftw	Forschungszentrum Telekommunikation Wien (Telecommunications Research Center Vienna)
GS	Guaranteed Service
H.323	ITU-T Standard for Packet based multimedia communications systems
IETF	Internet Engineering Task Force
IntServ	Integrated Services
IP	Internet Protocol
ITU-T	Telecommunications Standardization Sector of the International Telecommunications Union
kbit/s	Kilobits per second
MF	Multifield classification
MTU	Maximum Transfer Unit
PDB	Per-domain Behavior
PHB	Per-hop Behavior
PSTN	Public Switched Telephone Network
QoS	Quality of Service
RED	Random Early Detection
RFC	Request for Comments
RIO	RED In and Out
RSVP	Resource Reservation Protocol
SIP	Session Initiation Protocol
SLA	Service Level Agreement
TCA	Traffic Conditioner Agreement
TCP	Transport Control Protocol
ToS	Type of Service
WP	Work Package
VoIP	Voice over Internet Protocol
VPN	Virtual Privat Network
VW	Virtual Wire
WFQ	Weighted Fair Queuing
WRED	Weighted RED

- less loaded and therefore faster core routers (by limiting their queuing stage complexity);
- less states to be signaled, processed, and stored (cross-core characteristics can be expressed in terms of only a few aggregate behaviors).

Another important aspect to note in the DiffServ architecture is the existence of a Service Level Agreement (SLA), i.e. a contract between the subscriber and the provider. This is the essential link for pricing differentiated services. A subset of the SLA is the Traffic Conditioning Agreement (TCA), which specifies the detailed service parameters for each service level. The service classes may be defined via negotiation between both parties (subscriber and network provider).

Although others are possible, there are currently two standard PHBs defined that are building blocks for edge-to-edge services:

- Expedited Forwarding (EF): EF minimizes delay and jitter, assures bandwidth and provides the highest level of aggregated QoS. This service provides a virtual line PHB across a heterogeneous IP network, a precondition for applications with end-to-end guarantees. Any EF traffic that exceeds the traffic profile (which is defined by local policy) is discarded at the policer. The amount of aggregate traffic entering the network can never be greater than the amount of bandwidth capacity available to the EF service at any given link.
  An important example of the use of EF is the implementation of a low loss, low latency and low jitter Virtual Wire Per Domain Behavior (VW PDB) (previously also described as Premium service and Virtual Leased Lines) (*Jacobson, Nichols, Poduri, 1999*). The VW behavior is essentially indistinguishable from a dedicated circuit (Fig. 5) and can be used anywhere, it is desired to replace dedicated circuits with IP transport. Two steps have to be done: configuring the individual routers so that the aggregate has a well-defined minimum departure rate (by implementing the EF PHB) and conditioning the entire DS domain's aggregate (via policing and shaping) so that its arrival rate at any node is always less than that node's configured minimum departure rate. The VW PDB has two major attributes: an assured peak rate and a bounded jitter.
- Assured Forwarding (AF): The AF PHB group is a collection of PHBs that offer a high level of assurance that each packet will be delivered, as long as the traffic flow conforms to a given service profile, or TCA (*Heinanen, Baker, Weiss, Wroclawski, 1999*). Excess AF traffic is delivered with a lower probability than the traffic "within profile", which means it may be reduced in rank but not necessarily dropped. In contrast to EF (which has hard bandwidth and jitter characteristics), the AF group supports dynamic sharing of network resources and soft bandwidth and loss guarantees, which are useful for traffic with looser QoS requirements. AF consists of four classes and three possible drop precedences within each class (resulting in a total of twelve DSCPs). Each AF class is allocated a certain level of bandwidth and buffer space in each DiffServ node. Admission filtering must control the amount of traffic admitted at each AF level and may include traffic shaping, selective packet discard (see Sect. 3.4), and alteration of any marked discard preferences.

An example for creating a PDB using the AF PHB is the "Assured Rate Per-Domain Behavior for Differentiated Services" (*Seddigh, Nandy, Heinanen, 2001*). This PDB is useful for carrying traffic aggregates that require rate assurance but do not require delay and jitter bounds. The traffic aggregate will also have the opportunity to obtain excess bandwidth beyond the assured rate.

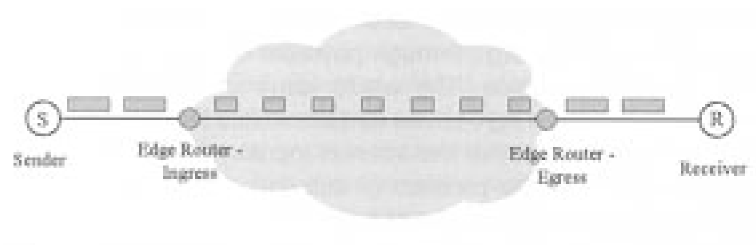

Fig. 5. Virtual Wire Concept

5. Voice over IP application example

After reviewing the important general QoS concepts, mechanisms and architectures in the previous sections of this paper, we now consider an example of a real application. The Voice over IP (VoIP) scenario described here is an important case for demonstrating how multimedia streams may be successfully transported within a QoS-enabled network. For further details on Voice over IP protocol usage (RTP/UDP/IP) and voice encoder parameters (G-Series) we refer to (*Collins, 2001*) and (*Hersent, Gurle, Petit, 2000*). In the rest of the section, we describe, analyze and generalize the standard example scenario as presented in (*Jacobson, Nichols, Poduri, 2000*).

Let us assume an enterprise which wants to use the Virtual Wire (VW) concept from Sect. 4.2 to provision a large scale, internal VoIP telephony system. Suppose that the internal links of the enterprise are all Ethernet (B = Mbit/s) having a frame length of $L = 1\,500$ bytes and that the network diameter is n hops (usual numbers for n ranging from 3 to 10). Typical telephone audio codecs deliver one packet every $\Delta = 20$ ms. At this codec rate, RTP encapsulated G.711 voice consists of packets of size $s = 200$ byte in a worst-case analysis, ignoring the effect of RTP header compression.

Suppose further that our enterprise's network usually is crowded with all sorts of data traffic (e-mails, WWW traffic, file transfer etc). In this case, traditional best effort could not support voice traffic sufficiently, whereas under VW, voice traffic is strictly prioritized over data traffic. We now derive a general expression for the number of parallel VoIP calls from one end of the network to the other that can be granted in a QoS-enabled (VW) network.

Let us first consider the limiting (best) case. If the entire network were dedicated to voice traffic only, each voice flow should be given the impression of having a dedicated circuit. Hence, the maximum tolerable jitter introduced by the network equals the inter-packet arrival time of $\Delta = 20$ ms. In a time window of 20 ms, a data volume of $V = B \cdot \Delta/8 = 25$ kbytes can be transmitted at a rate of $B = 10$ Mbit/s. 25 kbytes correspond to V/s = G.711 packets and thus a maximum of $N_{max} = 125$ calls which could be handled if the net were carrying only G.711 encoded telephony traffic.

In a second step, now we admit all types of traffic to the network. Assuming peak rate policing (token bucket size equals a few packets) and proper admission control the aggregate EF arrival rate over any time-interval longer than the packet service time on the virtual wire (20 ms in our case) is lower or equal than the service rate at the EF queue. This property holds for each hop (see Sect. 4.2). Thus a voice packet cannot be inserted behind another voice packet in the EF queue, i.e. VW jitter due to interference from other voice traffic equals zero.

Now we consider the maximal jitter of a voice packet caused by interference from non-voice traffic. In the worst case, at each hop a non-voice packet of length $L = 1\,500$ bytes has just started to be served by a non-preemptive packet scheduler at the arrival of our voice packet. Hence, the jitter is increased at this paraccess to an applied service level. Access can be established in a number of ways e.g., through payment of fees or ranking of the user's importance. The whole issue of establishing and monitoring a user's right to use certain service levels opens up a variety of questions that the Internet industry is only beginning to address. First, the problem of authentication – proving that the entity currently using the network is the claimed user, either during admission negotiations or subsequent traffic transmission. Second, the question of pricing and billing to optimize user satisfaction and network owner revenue. Billing is even of interest to enterprise networks, where it may be used to provide additional granularity of usage control beyond the corporate status of a user.

4. Current QoS architectures

4.1 Integrated Service (IntServ)

The IntServ architecture was proposed with the background that the basic best effort service needed some levels of modification in order to provide customized support for different service classes. For example a VoIP service implies some bounds on delay, loss and jitter, while data transfer requires bounds on available capacity. With the focus on supporting such individual applications, explicit a priori signaling of each flow's requirements were made mandatory. For the first time in packet networking the concept of resource reservation was added to the Internet service model. In a certain way IntServ and ATM have some properties in common, which makes IntServ and ATM mapping possible. To fulfill the role of the signaling protocol for IntServ the Resource Reservation Protocol (RSVP) was developed (*Braden et al., 1997; Wroclawski, 1997*).

To date, IntServ has developed support for two classes of services:

- Guaranteed Service (GS): This service class approximates as close as possible the behavior of a dedicated circuit. GS allows the receiver of a flow to specify the maximum allowable end-to-end latency and provides a firm bound on this adjusted latency. In addition it ensures bandwidth availability (*Shenker, Partridge, Guerin, 1997*). Thus GS is recommended for delay-intolerant applications, such as real-time applications, audio and video streams.
- Controlled Load (CL): This service attempts to provide end-to-end traffic behavior that closely approximates traditional best-effort services within the environmental parameters of unloaded or lightly utilized network conditions. If a router cannot provide such performance, it will reject QoS request. Hence, it is better than best effort, but cannot provide the strictly bounded service that GS promises. CL cannot make any latency guarantees, because it does not use specific values for control parameters that include information about delay or loss. But it implies that the requested bandwidth will be available at low error rates and reasonable latency. CL service is recommended for delay-tolerant applications. It provides a less-than-reliable delay bound. If the requested resources fall outside the bounds of what is available, the traffic flow may experience some small additional amount of induced delay or possibly a slightly higher rate of dropped packets. The degree at which traffic may be dropped or delayed should be low enough for an adaptive real-time application to be transmitted without noticeable degradation.

The disadvantages of IntServ are the complexity and overhead that makes it not appropriate for the Internet. If a router has to store information about individual flows, as at IntServ, the scalability suffers. Further, to allow network providers to offer different prices for different classes is not easily resolved with RSVP. But the deployment of IntServ is broadly accepted in combination with other QoS architectures such as ATM or the following explained Differentiated Services. In the combination with ATM, the services CL and GS of IntServ has to be mapped to the ATM services e.g. to establish an IP connection over ATM backbone. In combination with Differentiated Services there are many scenarios where IntServ mechanisms operate as access, while the Differentiated Services mechanisms operates in core routers.

4.2 Differentiated Services (DiffServ)

The IETF DiffServ architecture (*Blake et al., 1998*) emerged in 1997 as a complement to IntServ. A number of people argued successfully that a solution with simpler differentiation of traffic as proposed in IntServ was needed. The first specifications characterized differentiated services based on the assumption that IP packets should be marked with relative priority. This concept proved to have the desired scaling properties. Instead of responding to individual flows, DiffServ is designed to offer service to aggregates of classes, where a number of individual flows are grouped (aggregated). A single flow receives the same handling as all of the other flows in its aggregate. No per-flow state needs to be maintained in the routers, neither is there an explicit connection setup phase.

Traffic entering a DiffServ network at an edge router runs through the following components: classifier, admission filter, marker. Multifield classification (MF) and marking at the networks edge allow a wide variety of traffic to be mapped into a smaller subset of behaviors provided by the network's core. The admission filter examines and shapes the traffic to meet the requirements associated with the classification. Subsequently, the DS field (DiffServ field) in the IP header is marked depending on the particular aggregate to which this packet is assigned. The DS field is the byte for class of service categorization in an IP packet used in the DiffServ environment. This byte occupies the place formerly used by the TOS-byte. The TOS-byte (Type of Service) was (and still is where no DiffServ is implemented) the packet precedence and traffic classification byte in IPv4. Six bits of the DS byte are allocated to Differentiated Services Code Points (DSCP).

The behavior to which routers are triggered by the DSCP is called Per-Hop Behavior (PHB). The packet-handling of core routers solely depends on the DSCP carried in every packet. DiffServ nodes which operate with a consistent set of differentiated services policies administered in a coordinated fashion and with a set of PHB groups, are called to be in a DiffServ Domain (*Blake et al., 1998; Nichols, Carpenter, 2001*). A Per-Domain Behavior (PDB) is an expected treatment of packets in an end-to-end connection of a DiffServ Domain. A particular PHB (or, list of PHBs) and traffic conditioning requirements are associated with each PDB.

The reasons for the scalability of DiffServ are as follows:

- Complex decision making (e.g., multifield classification, conditioning of traffic, assignment of a DSCP value) is pushed to the edges.
- Useful edge-to-edge services are built from a restricted set of behaviors in core routers. The core routers have to respond to the service marking contained in the DSCP, and to give a service response that is consistent with this marking. No further classification is performed in the core network.
- Subsequent PHBs are triggered within the network by the DS field value.

Focused on the backbone of a service provider, where router performance almost always is a bottleneck and hundreds of thousands of flows exist, two key benefits are expected from DiffServ:

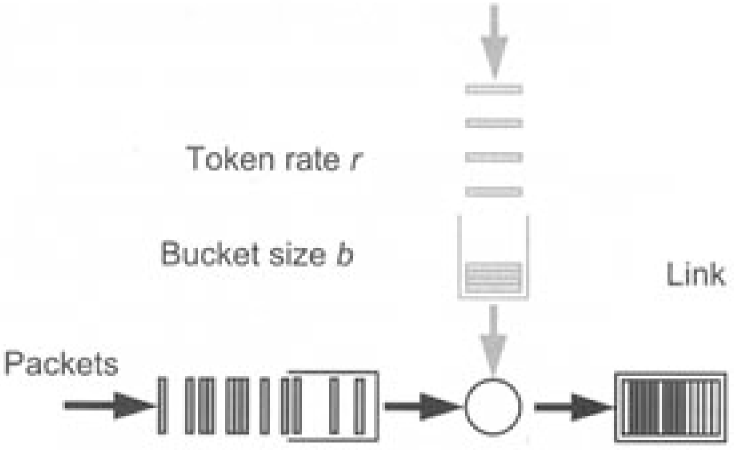

Fig. 4. Token bucket

packets that may arrive during a well defined time interval. This allowable behavior is the traffic profile. If a flow's rate conforms to the traffic profile, its packets are marked as "in-profile". If the flow's rate exceeds the limit specified in the traffic profile the excess traffic is marked as out-of-profile. Dependent on the traffic class, out-of-profile packets are either discarded or forwarded to the next hop router.

Traffic policing is generally performed by a token bucket mechanism (Fig. 4). Tokens are added to a bucket at some fixed rate of r (tokens per second) and are removed from the bucket whenever a packet arrives. The bucket has a defined depth b. Thus, if the bucket is full and a burst of length $c > b$ arrives then $(c-b)$ packets will be marked as out-of-profile.

3.3 Traffic shaping

Traffic shaping provides a mechanism at the sender side to smooth traffic bursts and thereby make the traffic confirm to the traffic profile. There exist two functionalities. The first one is used to smooth traffic bursts, such that all traffic is upper bounded by a peak rate (i.e. bursts are cut completely). The second is used to enforce a long-term average traffic rate and permits some degree of burstiness. Similar to traffic policing, shaping can be performed with a token bucket mechanism.

3.4 Queue management

Queue management decides when to drop packets based on measurements of the queue length at a congested router output port (as roughly defined in *Braden et al., 1998*). Thus queue management interacts closely with end-to-end congestion control. The goals are to achieve high throughput and low delay. The buffer space in the network is designed to absorb short-term data bursts rather than be continuously occupied. Limiting the queue size can help to reduce the packet delay bound. Besides generation of congestion feedback, queue management mechanisms have the task of preferentially dropping or marking out-of-profile packets against in-profile packets in order to enable low drop rates for in-profile packets in times of congestion (*Cisco Web pages, Clark*).

In the literature often both drop and scheduling mechanisms are referred to the same chapter. They are closely related, so that e.g., a specific dropping mechanism gives best results with only one specific scheduling mechanism in combination.

The traditional queuing management mechanism in Internet is tail-drop, implemented in combination with FCFS (First Come First Served) scheduling. Tail drop means simply that, if the output port buffer is full, arriving packets are discarded. This has several disadvantages concerning fairness and resource utilization. A refinement is gained through the methods of Active Queue Management (AQM), which detect the initial stages of congestion and start dropping packets early. The RED (Random Early Detection) AQM algorithm (*Floyd, Jacobson, 1993*) consists of two main parts: estimation of the average queue size and the calculation of a drop probability as a function of the average queue size. The average queue size is estimated using a simple exponentially weighted moving average. The probability of dropping increases linearly as the estimated average queue size grows. Weighted RED (WRED) (*Cisco Web pages*) and "RED In and Out" (RIO) (*Clark*) are extensions of RED to support preferential dropping of out-of-profile against in-profile packets in addition to the benefits concerning detection of incipient congestion inherent to active queue management.

3.5 Scheduling

Scheduling dictates the temporal characteristics of packet departures from each queuetypically at the output interface towards the next router. Because a packet's traffic class dictates which queue it is placed in, the scheduler is the ultimate enforcer of relative priority, latency bounds, or bandwidth allocation between different classes of traffic. A scheduler can establish minimum available bandwidths for a particular class by ensuring that packets are regularly pulled from the queue associated with that class. A scheduler can also provide rate shaping (imposing a maximum allowed bandwidth for a particular class) by limiting the frequency by which a queue is serviced.

The previously mentioned FCFS is not applicative for QoS networks, as it cannot allow special handling of prioritized packets. Traditionally unresponsive flows (flows without congestion control) may unfairly consume the entire link bandwidth.

To address these limitations a number of scheduling disciplines have been developed to meet QoS scheduling requirements like distinguishing of traffic classes and complying with bandwidth guarantees:

- Priority scheduling provides a separate queue for each priority class. Basically, it is a multiple-queue FCFS scheduling discipline with higher priority queues being served first. Consequently, packets are scheduled from a particular priority queue in FCFS order only if all higher priority queues are empty. In such a model, the highest priority traffic receives minimal delay, but all other priority levels may experience resource starvation if the highest precedence traffic queue remains occupied. To ensure that all traffic receives some minimum level of service admission control has to be associated that restricts the amount of traffic for each priority level.
- Fair Queuing (*Demers, Keshav, Shenker, 1989*) maintains a separate queue for each traffic class and uniformly distributes link bandwidth among queues. Weighted Fair Queuing (WFQ) (*Parekh, Gallager, 1993*) extends fair queuing by allowing to assign weights to each individual queue enabling bandwidth distribution among queues different from the uniform distribution. Variants of WFQ have been proposed, where a trade-off is made between the complexity and delay bounds. In general WFQ ensures a minimum level of resource allocation to a specific service class. In combination with some admission filtering, the result will be bounded delay as well as guaranteed throughput.

3.6 Authentication, billing and pricing

Any networking technology that offers differentiation of service levels must also address the need to differentiate each user's right to use particular service levels. The danger always exists that users will try to attain a higher service classification than to which they are entitled. The problem is allowing a particular user

References

Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., Weiss, W. (1998): An architecture for Differentiated Services. Request for Comments 2475.

Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J., Zhang, L. (1998): Recommendations on queue management and congestion avoidance in the Internet. Request for Comments 2309.

Braden, R. (Ed.), Zhang, L., Berson, S., Herzog, S., Jamin, S. (1997): Resource ReSerVation Protocol (RSVP). Request for Comments 2205.

Cisco Web pages. http://www.cisco.com/warp/public/732/netflow/qos_ds.html/.

Clark, D.: Explicit allocation of best effort packet delivery service. http://www.ietf.org/html.charters/diffserv-charter.html/.

Collins, Daniel (2001): Carrier Grade Voice over IP. McGraw-Hill Professional Telecom.

Demers, A., Keshav, S., Shenker, S. (1989): Analysis and simulation of a fair queueing algorithm. Proceedings of ACM SIGCOMM.

Ferguson, P., Huston, G. (1998): Quality of Service. Wiley.

Floyd, S., Jacobson, V. (1993): Random early detection gateways for congestion avoidance., IEEE/ACM Transactions on Networking.

Greenville Armitage (2000): Quality of Service in IP networks. Macmillan.

Heinanen, J., Baker, F., Weiss, W., Wroclawski, J. (1999): Assured forwarding PHB Group. RFC 2597.

Hersent, O., Gurle, D., Petit, J.-P. (2000): IP telephony – packet-based multimedia communications systems. Addison-Wesley.

Jacobson, V. (1988): Congestion avoidance and control. Proceedings of SIGCOMM '88; Palo Alto, CA.

Jacobson, V., Nichols, K., Poduri, K. (1999): An expedited forwarding PHB. RFC 2598.

Jacobson, V., Nichols, K., Poduri, K. (2000): The 'Virtual Wire' Per-Domain Behavior. Internet Draft. IETF.

Kilkki, K. (1999): Differentiated services for the Internet. Macmillan.

Nichols, K., Carpenter, B. (2001): Definition of differentiated services per domain behaviors and rules for their specification. RFC 3086.

Parekh, A. K., Gallager, R. G. (1993): A generalized processor sharing approach to flow control in integrated services networks: The single-node case. IEEE/ACM Transactions on Networking.

Reichl, Peter, Stiller, Burkhard (2001): Edge pricing in space and time: theoretical and practical aspects of the cumulus pricing scheme. Proceedings of the 17th International Teletraffic Congress ITC-17, Salvador da Bahia, Brazil.

Seddigh, N., Nandy, B., Heinanen, J. (2001): An Assured Rate Per-Domain Behaviour for Differentiated Services. Internet Draft.

Shenker, S., Partridge, C., Guerin, R. (1997): Specification of guaranteed quality of service. RFC 2212.

Siegel, Eric D. (2000): Designing Quality of Service. Wiley.

Stevens, W. Richard (1994): TCP/IP illustrated volume 1. Addison-Wesley.

Uless Black (2000): QoS in wide area networks. Prentice Hall.

Wroclawski, J. (1997): The use of RSVP with IETF integrated services. Request for Comments 2210.

Ziegler, T., Brandauer, C., Fdida, S. (2001): A quantitative model of RED with TCP traffic. Proceedings of IEEE/ACM/IFIP IWQoS 2001, Karlsruhe. ■