TOWARDS A SCALABLE SYSTEM FOR PER-FLOW CHARGING IN THE INTERNET

Gabriel Dermler[1], Manuel Günter[2], Torsten Braun[2], Burkhard Stiller[3]
[1]IBM Research Laboratory, Zürich, Switzerland
[2]Institute of Computer Science and Applied Mathematics IAM, University of Bern, Switzerland
[3]Computer Engineering and Networks Laboratory TIK, Swiss Federal Institute of Technology, Switzerland
E-Mail: [1]gde@zurich.ibm.com, [2]{mguenter, braun}@iam.unibe.ch, [3]stiller@tik.ee.ethz.ch

Keywords: Resource Management, Network, Arrival generation, Model evaluation

ABSTRACT

The provision of guaranteed Quality of Service (QoS) in the Internet requires appropriate system support for both resource allocation and charging. Differentiated Services is an approach for the former which targets a high level of scalability. The inclusion of flow-based charging characteristics, such as QoS, extent of service usage and traffic destination dependencies, into Differentiated Service models requires specific system components including destination related price tables, traffic counters and specific control schemes for inter-provider service agreements. In this paper, we describe and evaluate these components with respect to their spatial and temporal scalability. For the latter, an evaluation model is developed based on simulations and simulation results are provided indicating performance trade-offs.

INTRODUCTION

Important distributed applications, such as tele-conferencing, IP telephony or on-demand media delivery require QoS support at the network level. While the Internet has been tremendously successful at supporting communication with no or relatively low QoS semantics, the incorporation of stringent QoS concepts has not happened so far on a significant scale.

One reason for this is that there is no conclusion as to which QoS framework to support. Such frameworks are important in order to define end-to-end services across multiple Internet Service Providers (ISPs) in the public Internet. The early concept of Integrated Services (IntServ) considered QoS support on an end-to-end basis (Braden et. al. 1994). However, its per-flow service model proved to have severe disadvantages. On one hand, IntServ is able to provide differentiated QoS support to applications on a per-flow basis. On the other hand, maintaining this model throughout the network implies per-flow state overhead which was shown to lead to severe scalability problems especially in the core parts of the network.

In response to such deficiencies, Differentiated Services (DiffServ) have been put forward (Blake et. al. 1998). DiffServ assumes a network model consisting of interconnected domains, where domains can be operated by different network service providers. To improve scalability within a domain, QoS support is provided in terms of a limited number of QoS classes. Between interconnected domains, Service Level Agreements (SLAs) are foreseen as a means to regulate the traffic exchanged and the service provided. Although SLAs and their enforcement are essential for preventing network domains from congestion, the issue of how to define and handle SLAs is at an early stage of consideration. In principle, SLAs can be defined at various aggregation (and hence scalability) levels including the cases of per-flow and per QoS-class agreements (Bernet et. al. 1999).

The question of how to charge for Internet usage is closely related to the introduction of QoS support. Obviously, as soon as such support becomes available, network users will tend to make use of the best available service only, unless they have to pay more for a better service. Assuming a DiffServ type of QoS provision, charging for QoS can be related to the type of SLAs employed. For instance, if class-based SLAs were to be used, charging could be based on the selected QoS class. If SLAs are employed on a per-flow basis, charge calculation can in addition reflect the communication path including the destination implied. Such charging allows for an accurate determination of actual cost of communication on a fine granularity level of usage. As a close linkage between cost and applied charge is a natural outcome in competitive environments, these properties of per-flow charging are desirable for the Internet as well (Stiller et al 1999).

The charging system presented in this paper attempts to combine the benefits offered by per-flow charging with the increased scalability of QoS support as promised by DiffServ. The proposed system includes the ability to charge based on the used QoS class, requested bandwidth as well as the destination of communication. Its realization is based on the DiffServ network model mentioned above, where ISPs offer inter-provider SLAs as chargeable services. In order to achieve higher scalability, an SLA control scheme is proposed which detaches the set-up and adaptation of SLAs from the set-up and termination of individual flows. The scheme can adequately support QoS as well as the charging aspects mentioned above. However, it implies a certain level of resource overbooking, thus introducing a trade-off between increased scalability of SLA handling and the extent of resource underutilization. An initial evaluation of this trade-off constitutes a second major part of this paper.

The remainder of the paper is structured as follows. In the next section, the charging system is introduced in terms of SLAs and pricing. Then, an SLA control scheme is presented which allows for increased scalability of SLA treatment. A simulation model is described allowing to assess the trade-off between resource overbooking and SLA signalling overhead. Simulation results are provided indicating the nature of this trade-off. Finally, a short review of related work is provided along with conclusions of the work done.

A DIFFSERV BASED CHARGING SYSTEM

The figure below depicts the typical setting of a DiffServ defined network environment. Multiple network domains, operated by possibly different ISPs, have to be crossed by flows exchanged between applications running on hosts. Within each network domain, a number of network services is assumed to be implemented including the traditional best-effort service and at least one additional service for QoS support. For the latter, several different approaches have been proposed including the Olympic, Assured and Premium type of services (Baumgartner et al. 1998). Throughout this paper, we tacitly relate our work to the Premium Service which is assumed to be associated with low delay and loss bounds for traffic delivery.

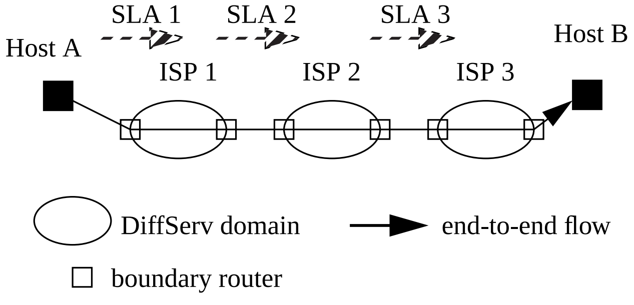

Figure 1: DiffServ network model

The issue of how QoS based services can be implemented within a domain has been considered elsewhere (Xiaio et al. 1999, Braun et al.1999). In short, for each domain a so-called bandwidth broker is foreseen which is responsible for admitting traffic to entering the domain through its boundary routers. Based on knowledge about available resources within the domain, the broker is assumed to be able to decide on new traffic requests such that the QoS service characteristics (bounded loss and delay for Premium Service) are not violated. Admittance of new traffic requests is handled in terms of Service Level Agreements. An ISP requests an SLA from a neighbor ISP, if it wants to send traffic to it. The ISP offering the SLA has to agree with the requesting ISP on the amount of traffic (bandwidth) to be sent and the type of service to be provided.

In order to ensure end-to-end QoS support, SLAs have to exist along the path of ISPs connecting a flow's source and destination. The SLA semantics need to imply a service coverage extending from the contracting ISPs to the destination end-points of the traffic covered. For instance, in Figure 1, the SLA between ISP1 and ISP2 needs to give service assurance to ISP1 for QoS (e.g. delay) occurring between the ingress point of ISP1 and the egress point of ISP3 (connecting to host B). Naturally, ISP2 itself needs an SLA with ISP3 in order to judge service availability up to the desired destination end-point. This "nesting" of SLAs is a prerequisite to the provision of end-to-end QoS.

Introducing charging into the described environment requires a number of additional considerations. As the SLA is the unit of service commitment between two ISPs, it provides a natural context for defining the charging to be applied. In this paper, we assume a straight-forward charging semantics for an SLA. We assume that unit bandwidth prices can be derived based on the QoS class, the bandwidth amount agreed on as well as the targeted destination. The final charge for an SLA is obtained by using this price and the time interval during which the SLA is maintained.

In order to enable such charging differentiation, an ISP needs to maintain an SLA for each serviced destination. If the number of destinations is too large, the ISP may face scalability problems similar to the ones implied by IntServ, as the only limit is given by the number of Internet hosts. The situation can be improved by aggregating sets of destinations and servicing them using a common SLA. Various approaches may be conceivable, for instance by considering all destinations in a a region as being subject to the same unit price. However, such an approach assumes that all ISPs in a region can be reached and traversed at approximately the same price level.

We propose an alternative approach based on an IP address aggregation method which is already in use by the BGP (Border gateway protocol) for inter-domain routing in the Internet (Rechter and Li 1995). Briefly, routing across domains is performed based on the network address part of the IP address only, identifying the destination network to which the destination host is connected. At border routers of domains, routing entries must be kept in principle for each occurring destination network in the Internet, but not for every occurring host IP address. Border router of (core) ISPs typically hold entries of several 10.000 networks without facing scalability problems.

By augmenting the routing tables with a price entry a destination dependent charging scheme can be enabled. Each ISP keeps track of the price for every destination network. In particular, an access ISP is able to provide such information when queried by an end user for a new flow. Pricing information may consist of a uniform bandwidth unit price or prices for various bandwidth amounts. Such price tables may be rather static in nature, or be updated on a regular basis, for instance every week and be performed at a time when update traffic does not cause resource congestion.

SLA CONTROL AND SIGNALLING

Given the mentioned address aggregation, each ISP needs to keep track of the traffic directed to every possible destination network. We propose a mechanism, termed Destination Related Reservation Tracking (DRT) which is to ensure that each ISP is aware of the traffic amount for which QoS support has been committed. More precisely, each ISP is expected to have a counter for each destination network mirroring the overall bandwidth admitted towards that network. This approach contains the SLA scalability problem in terms of storage: instead of keeping track of per-flow state, per destination network state suffices. Along the temporal dimension the situation appears to be unchanged: each ISP is expected to capture each flow set-up/termination indication, establish the corresponding destination network, upgrade its SLA with the downstream ISP for the additional bandwidth and update the corresponding traffic counter in its combined price/admitted traffic table.

We foresee a second mechanism, termed Inter-Provider SLA Control, in order to ensure that an ISP does not have to adjust its SLAs towards downstream ISPs too frequently. Two aspects are underlying the approach. One is the fact that considering just one SLA between two neighbor ISPs aggregates traffic across all possible destinations. Although the aggregate components may change (i.e. flows are terminated and new flows with possibly different destinations are set up), such changes may to a large degree be averaged out when considering the traffic aggregate. Furthermore, the amount of the traffic aggregate may have both rather large and small fluctuations. At least the portion of small fluctuations can be prevented from triggering SLA updates, supposed the bandwidth contracted through the SLA is kept large enough.

The control scheme we propose in effect decouples SLA upgrades from the DRT process. At each point in time an ISP is aware (through DRT) of the admitted traffic towards a neighbor ISP. Instead of contracting for this amount only, the ISP is assumed to "overbuy" bandwidth in expectation of traffic fluctuation which it would like to support without having to readjust the SLA. More precisely, we assume that an ISP adopts the following behavior based on an overbuying measure d: whenever the bandwidth contracted through the SLA is fully used up, the ISP attempts to increase the bandwidth by at least d percent. Conversely, whenever flows are terminated, the contracted bandwidth is only decreased, if the amount of contracted, but unused bandwidth exceeds d %.

As detailed in the next section, this approach leads to contracted bandwidth which on average will be approximately utilized to 100-d/2 percent only. As charging is based on the contracted amount, the approach leads to both positive and negative effects. On one side, it decreases the frequency at which SLAs need to be updated. On the other hand, it implies unused contracted traffic and a corresponding an economic loss. We analyze this trade-off in the next section.

The consideration of only one SLA between two ISPs has another implication with respect to the price to be applied. The exact price for a contracted traffic aggregate is a function of the distribution of the destination networks of the traffic. In principle, the price for an SLA needs to change whenever a new flow is set-up respectively terminated. Adhering to our approach to avoid per-flow signalling overhead whenever possible, we assume a sampling of the traffic distribution at a time interval which is significantly large than the interval of flow changes. For instance, an ISP could check every 5 min the current distribution for a contracting neighbor ISP and announce a fixed price for the next 5 minutes based on this distribution. Obviously, such an approach may again have an economic influence on the charges an ISP is able to collect. However, the study of this aspect is outside the scope of this paper.

Inter-provider interaction is decomposed by our approach into three different signalling levels.On one level, per-flow signalling takes places in order to admit new flows and update DRT counters of ISPs. On another level, SLA updates take place based on update policies. On a third level, price announcements are distributed at regular intervals.

While the approach does not fully eliminate per-flow related overhead, we see its benefit in the fact that it detaches business related interactions from pure admission signalling. As SLAs imply a contractual relationship, higher processing requirements due to negotiation and security requirements, have to be assumed. These aspects render a low SLA update frequency a primary goal.

EVALUATION MODEL

Initial simulations have been carried out in order to evaluate the behavior of the SLA control scheme. They are based on the sample network depicted below, where each node denotes an ISP accepting and sending signal messages from/to neighbor ISPs. We use a combination of hierarchical and meshed interconnections in order to approximate the structure occurring in the Internet. Access networks to which hosts can connect form the leaves of the structure. Simulations have been carried out using various network sizes and number of levels in the hierarchy. The results indicated in this paper are based on a 3-level network consisting of 12 backbone and 24 host networks.

Our study concerns traffic associated with one QoS service class only. Access networks are assumed to be the sources resp. sinks of traffic. Decisions on new flow generation resp. flow termination are done based on simulation rounds. Within each round, an access network checks a number of times (maxNew) whether a new flow is to be generated. Each generation is based on probability probNew. Per round at most maxNew flows can be generated and the average generation is given by newMax*probNew. Existing flows are associated with a termination probability probTerm within each round. Thus, the average life-time of a flows amounts to 1/probTerm. With these parameters, the number of flows in the considered network averages (in a balanced generation/termination state) to numberOfAccessNW*maxNew *probNew/ probTerm. For the presented results we assume: newMax = 10, probNew = 0.1, probTerm = 0.1. This implies an average number of 240 flows in the network.

For each simulation, a certain bandwidth capacity is assumed for the links interconnecting the ISP domains. Within the domains, resources are assumed to be abundant. Across simulations, link capacity was varied in order to create resource bottlenecks of various extent (see below). Each simulation consists of a number of rounds sufficiently large to get beyond the state where flow generation and termination reach balance. Initially, the network is considered empty. In each round, flow generation and termination are simulated using the parameters mentioned above. When a flow is created, end-to-end set-up signalling is performed. If existing SLAs along the path suffice, the flow is set up and the DRT counters of the involved ISPs are updated. If contracted bandwidth does not suffice at least at one ISP, the flow is not set up.

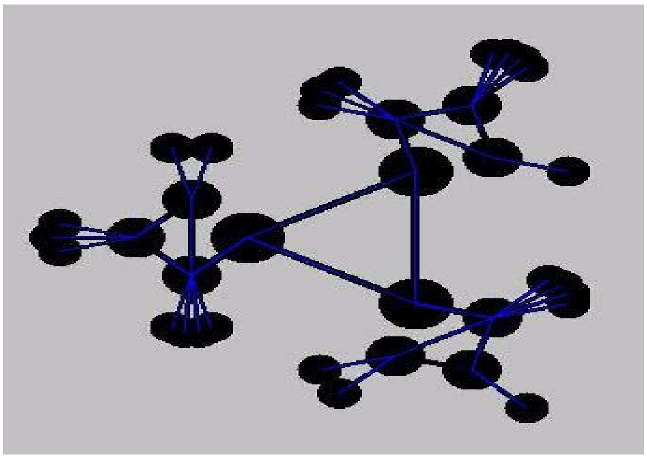

Figure 2: Employed ISP network

Each ISP maintains two SLAs with each of its neighbors, one for the downstream and one for the upstream direction. An SLA is managed by the ISP using it for sending the traffic. Upon receiving a flow set-up or termination event in a simulation round, an ISP checks whether it needs to adjust the contracted capacity in downstream direction. For this purpose, it employs an SLA update policy based on two threshold values, an increase_threshold (*it*) resp. descrease_threshhold (*dt*). Whenever a new flow in a round causes the SLA utilization to surpass *it* percent, the ISP requests an increase in contracted bandwidth. Whenever a flow termination in a round causes the SLA utilization to drop below *dt*. the ISP requests a decrease. In both cases, the update is assumed to be performed in such a way, that the new amount of contracted bandwidth leads to an SLA utilization which is exactly between *it* and *dt*:

$$contractedBW = \frac{usedBW}{0.5it + 0.5dt}$$

In our experiments, we set *it* to 1.0, meaning that new flows trigger an SLA update only if the bandwidth is fully utilized. By varying *dt*, both the SLA update frequency and the amount of overbooked resources are determined. With the presented updated policy, the overprovisioned bandwidth approximately averages to: $\frac{(it - dt)contractedBW}{2}$.

A special case is given, if *it*=*dt*=1.0. In this case, every flow generation resp. termination leads to an SLA update. Throughout all simulations we used this IntServ like SLA treatment as a reference scenario.

EVALUATION RESULTS

In the first simulation, we assume that capacity is abundant throughout the network of Figure 2, such that every flow can be admitted. We are interested in how overprovisioning affects the number of SLA updates. We use the parameter *dt* with values 1.0, 0.8, 0.6 and 0.4 to indicate the former in Figure 3. Each depicted curve indicates the average reserved bandwidth per SLA in the network. Note that the curve for dt=1.0 indicates the case that no resources are overbooked.

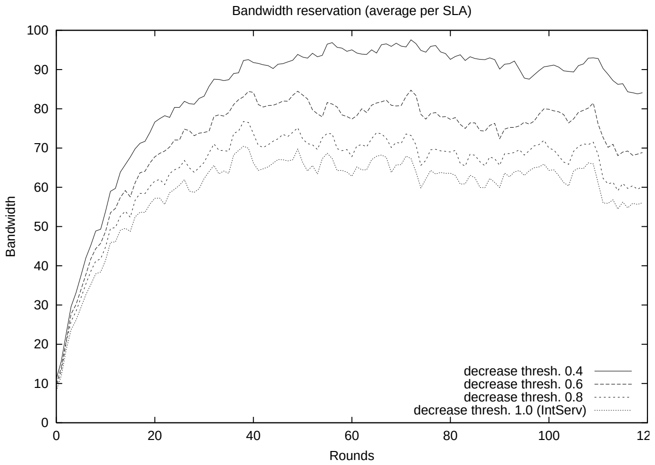

Figure 3

In Figure 4, the number of SLA updates is shown for the same *dt* values. Obviously there is a significant reduction of updates for *dt* < 1.0, more pronounced for *dt* values of 0.6 and 0.4. Resource overbooking reaches for these cases 20% and 30%, respectively. The implication of the latter is correspondingly higher service cost which for each backbone ISP is compensated as both the price to be paid to downstream ISPs is higher as is the revenue expected from upstream ISPs. Of course, the cost is ultimately reflected in the price which the (sending side) hosts have to pay.

However, overprovisiong does not necessarily accurately reflect the loss in possible revenue due to rejected flows. In order to reveal this aspect, we reduced the capacity of the network links and introduced resource bottlenecks. While most of the traffic still fits into the network, some flows have to be rejected. The following table shows the number of rejects during a 60 round simulation. Again, the *dt*=1.0 case shows how many rejections are due to true capacity limitations (an not to SLA contracting inefficiency):

dt=1.0	rejected : 79	loss: 0
dt=0.8	rejected: 90	loss: 11
dt=0.6	rejected: 107	loss: 28
dt=0.4	rejected: 132	loss: 53

During the 60 rounds roughly 1400 flows are attempted to be set up. Of these, 79 are rejected in the reference case. Expressed in economic terms, the network is able to generate revenue from roughly 1300 flows during that period. If resources are overbooked trough SLAs, still around 1250 flows can be supported in all of the considered cases. I.e. the reduction in revenue generation is in the range of less than 5%. This is an interesting result, as it allows the interpretation that the extent of overprovisioning does not necessarily represent the loss in financial revenue and that the loss may in fact be lower than the overprovisioning rate. Part of the explanation of this difference is due to the fact that the main resource bottleneck was given in the highest level backbone of the network and that flow destinations were set up randomly across all possible destinations. However, a deeper understanding of the relationship of network characteristics and revenue loss needs additonal consideration.

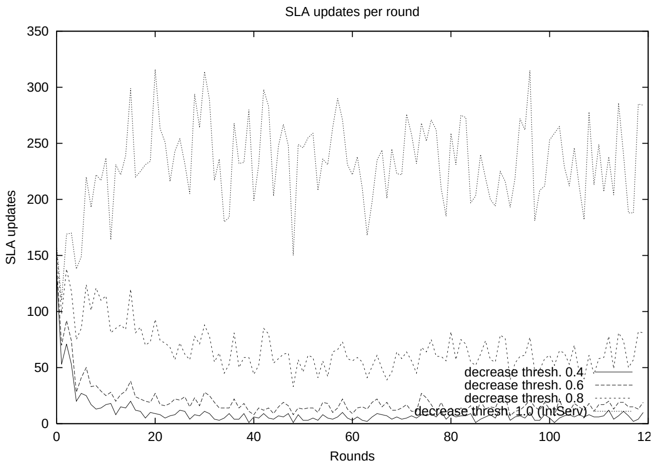

Figure 4

In the cases above, we assumed that an ISP is able to perform an SLA update whenever one is needed. However, this ability may be limited as well. To take this into account, we consider an update-wait time *w* given in number of rounds. If *w*=0, SLAs can be updated at any time. If *w*=1, after each update a wait time of one round is required before the next update can be performed. Again, this mechanism is tested with varying values of *dt*. If *dt* is large, then overprovisioning is low and many SLA updates are necessary. If we artificially limit the number of updates, new flows cannot be admitted, although resources would have been available (i.e. “contractable”). With *dt*=0.8 and *w*=2 we found the (unacceptable) behaviour that 60% of all new flows are rejected.

The situation is different for lower values of *dt*, e.g. 0.4. Here, *w*=2 still leads to acceptable results. As shown in the figure below, the flow rejection rate drops to a low stable level. However, during traffic surges, as during the initial phase of the simulation (with no SLAs initially), still high rejection rates can be seen. If no wait time for updates is assumed, the number of SLA updates is much higher (e.g. 134 instead of 35 in the first round) and the network becomes much more rapidly stable (in 20 instead of 50 rounds).

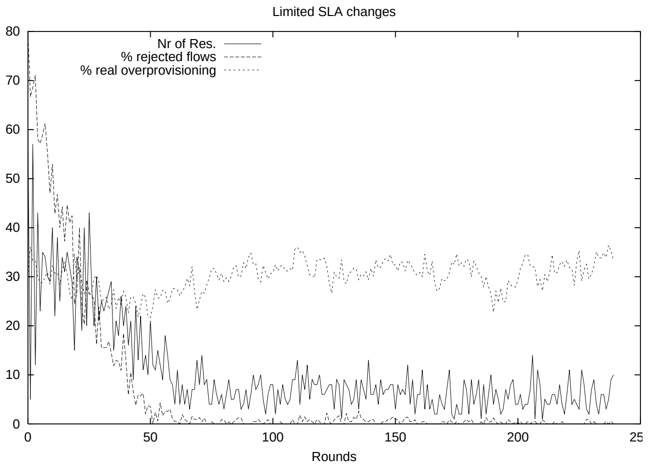

Figure 5

Related Work and Conclusions

The issue of inter-provider SLA set up has received less attention than the definition and implementation of service implementations inside DiffServ domains. The schemes which are closest to our approach relate to adaptive SLA adjustments based on measured cross-domain traffic (Terzis et al. 1999, Guenter et. al. 1999). However, these approaches cannot ensure QoS guarantees.

The address aggregation method mentioned in the paper was earlier used to determine an ISP path towards a destination based on the cost implied by that path (Fankhauser et al. 1999). The approach assumed that each ISP keeps track of SLAs towards all possible destination networks and did not consider the possibility of aggregating these SLAs into just one and reducing the extent of SLA updates.

In contrast, the approach investigated in this paper is focussed on the provision of guaranteed end-to-end QoS and aiming at the description of a charging system which is better scalable than an IntServ like approach. We see its main contributions in two areas. The approach separates the necessary signalling between ISP domains into several components. Flow signalling, SLA updates and the exchange of pricing information are shown to be supportable in a decoupled fashion on different time scales. Such separation is possible, even if destination dependent charging is assumed. In order to support the latter, the required system elements such as price tables, destination traffic counters as well as complex SLA update controls were introduced.

The simulation results indicate that resource overbooking on a fairly small scale (20 to 30%) already has the potential to significantly reduce the number of SLA updates. However, they also indicate that still a large number of updates is necessary, if traffic changes suddenly. If the number of such updates is limited, either a high number of flow rejections is implied or a high level of resource overbooking is necessary. In conclusion, the benefit of the proposed SLA update control closely depends on the type of assumed traffic generation. In cases where traffic is changing rather slowly both in amount and direction, the approach is of proven benefit. Finding out the exact limits of the approach requires additional work with respect to various traffic generation patterns.

References

Baumgartner, F.; Braun, T. and Habegger B. 1998. “Differentiated services: A new approach for quality of service in the Internet.” *In Proceedings of the IFIP TC-6 Eighth Conference on High Performance Networking HPN’98* (Vienna, Austria, September 21-25).

Bernet, Y.; Binder, J.; Blake, S.; Carlson, M.;and Carpenter, B. 1999. “A Framework for Differentiated Services.” *Internet Draft (work in progress).*

Blake, S.; Black, D.; Carlson, M.; Davies, E.; Wang, Z. and Weiss, Z. 1998. ”An Architecture for Differentiated Services.” *Request for Comments RFC 2475*. IETF.

Braden, R.; Clark, D. and Shenker, S. 1994. “Integrated Services in the Internet Architecture: An Overview” *Request for Comments RFC 1633* . IETF.

Braun, T.; Günter, M.; Kasumi, M. and Khalis, I. 1999. “Virtual Private Network Architecture.” *Public CATI project deliverable available at www.tik.ethz.ch/~cati.*

Fankhauser, G. and Schweikhart, D. 1999. “Trading of Service Level Agreements.”, *Public CATI project deliverable available at www.tik.ethz.ch/~cati.*

Günter, M. and Braun, T. 1999. “Evaluation of Bandwidth Broker Signalling” *In Proceedings of the International Conference on Network Protocols ICNP’99*. IEEE Computer Society.

Rechter, Y. and Li, T. 1995. “A Border Gateway Protocol (BGP-4).” *Request for Comments RFC 1771*. IETF.

Stiller, B.; Braun, T.; Günter, M. and Plattner, B. 1999: “Charging and Accounting Technology for the Internet.” *In Proceedings of the 5th European Conference on Multimedia Applications, Services, and Techniques (ECMAST'99)*. Springer Verlag, Berlin, Germany.

Terzis, A.; Wang, L.; Ogawa, J. and Zhang, L. 1999. “A Two-Tier Resource Management Model for the Internet”, *Global Internet Conference 99*. December 1999.

Xiao, X. and Ni, L. M. 1999. “Internet QoS: the Big Picture.” *IEEE Network*, vol. 13, number 2, March/April 1999.