Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Placi Flury, Peter Reichl,
Jan Gerke, Burkhard Stiller

Charging Considerations for Virtual Private DiffServ Networks

TIK-Report
Nr. 94, August 2000

Institut für Technische Informatik und Kommunikationsnetze
Computer Engineering and Networks Laboratory

Placi Flury, Peter Reichl, Jan Gerke, Burkhard Stiller:
Charging Considerations for Virtual Private DiffServ Networks
August 2000
Version 1.0
TIK-Report Nr. 94

Computer Engineering and Networks Laboratory,
Swiss Federal Institute of Technology (ETH) Zurich

Institut für Technische Informatik und Kommunikationsnetze,
Eidgenössische Technische Hochschule Zürich

Gloriastrasse 35, ETH-Zentrum, CH-8092 Zürich, Switzerland

Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

Charging Considerations for Virtual Private DiffServ Networks

ANAISOFT Deliverable WP3

Version 1.0
August 2000

P. Flury, P. Reichl, J. Gerke, B. Stiller

Computer Engineering and Networks Laboratory, TIK, ETH
Gloriastrasse 35, CH–8092 Zürich, Switzerland
E-Mail: [flury| reichl| gerke| stiller]@tik.ee.ethz.ch

Contents

1 Abstract

Temporary networks, i.e. virtual private networks (VPNs), founding on Internet technologies are basically fueled by either a congestion control or a quality of service (QoS) provisioning approach. The denominator of both approaches is the inevitable need for pricing mechanisms, due to cost recovery and system protection.
The differentiated services (DiffServ) framework allows to built up scalable and QoS reliable VPNs. The pricing scheme for virtual private DiffServ networks requires to be flexible, simple, understandable and thus enjoy the acceptance of the customers.
On the base of investigations on actual pricing models an adaptive flat rate pricing scheme is presented, which suits well for VPNs built on DiffServ.

2 Introduction

Globalization stimulates companies to reduce costs and to augment sales. The ANAISOFT project investigates on workflow systems, agent platforms, automated workflow trading systems and communication infrastructures, with the objective to enable companies to participate on a system of Federations of Workflow Trading Systems (FWTS). In such a FWTS enterprises publicate in– and outsourcing commissions and offers. On this foundation, companies find the suitable partners for project achievement. The interoperation suggests the deployment of inter–organizational workflows, of workflow managing systems, of workflow trading systems and bidding protocols, of intelligent and mobile agents and of the interconnection of the involved enterprises or enterprise sections by temporary networks.
This deliverable deals with the interconnection of enterprises by temporary networks. The attention is turned to spontaneous interconnection scenarios, where the period from set up to release time can be very short, i.e. temporary networks of short lifetime. Such temporary networks need furthermore to support charging and accounting mechanisms, due to cost recovery and resource protection.
Establishing temporary networks between different companies or between geographically distributed enterprise branches is generally done by *virtual private networks* (VPN). VPNs can be divided in two categories, in *traditional VPNs* and *Internet VPNs*.
Technically a traditional VPN is set up by leasing communication links of fix bandwidths, e.g. some T1 links. From an economical point of view traditional VPNs give raise to large amounts of fixed costs, due to installation and monthly fees, and of variable costs due to mileage charges.
Internet VPNs instead, are technically based on the use of the current Internet. Compared to traditional VPNs this has a considerable impact on cost

reduction. Admittedly Internet VPNs are still a field of intense investigation covering security, *quality of service* (QoS) and the resulting charging matters.
The following sections analyze Internet VPNs, which use the *differentiated services* (DiffServ) technology to satisfy the requirements of interconnected enterprises in terms of QoS. The focus is set on the mechanisms of charging and accounting the offered QoS, with the purpose of optimal regulation of resource usage and with the purpose to guarantee the revenues of the involved ISPs.

3 Internet Virtual Private Network (VPN)

Due to different setup objectives, VPNs can be classified in dial–up and LAN–LAN VPNs. What allies both is the requirement and stringent necessity for security, where security encompasses authentication, authorization, confidentiality and data integrity. The techniques to provide security are not subject of this deliverable. The following examination on dial–up and LAN–LAN VPNs concentrates explicitly on Internet VPNs.
Dial–up VPNs have gain in importance given the augmented possibilities in nomading computing. Until recently companies had to maintain costly modem–banks to enable remote access and branch to office connectivity. Meanwhile Internet VPNs offer local connection points, so called points–of–presence (POPs), where users connect on. Connectivity, transport and security details are handled by the customer's ISP. In terms of QoS, dial-up VPNs usually don't have exorbitant requirements. The actions of mobile customers typically enfold information retrieval/upload, like e-mailing, FTP as well as access on internal documents. Applications with high QoS commitments, like IP-telephony and video-conferencing, although exotic for some companies, will gain on importance and require likewise to be supplied.
The characteristics of dial–up VPNs between branches and offices are quite similar to these of mobile customers and their company. They may differ in the traffic amount and in the duration of the connections, e.g. semi-persistent connections. However, there is a boundary, where the costs and the expenditures of a dial-up interconnection of branches and offices becomes critical and where LAN–LAN VPNs suit better.
The objective of Internet LAN–LAN VPNs targets at an technically equivalent replacement of traditional leased line VPNs and simultaneously at economical advantages for companies and the ISPs. Technical equivalence presumes full supply of the benefits of leased lines, i.e. especially of QoS characteristics. The economical advantage subsists on the appliance of the Internet. Communication lines are shared and are thus used more efficiently as leased lines. The ISP can sell more resources, whereas the company only needs to pay for the resources it uses.

4 QoS, Congestion, and the Internet

In the last decade huge investigation efforts on reserving resources in the Internet have been made. Their main offsprings are called IntServ, which describes the way of providing integrated services and which lacks of scalability, and DiffServ, which tries to avoid the drawbacks of IntServ and which is therefore still a hot topic of research.
However, today's only systems in widespread use allowing resource reservation and which offer hard QoS guarantees, are typically based on connection oriented network technology, e.g. telephony (POT) and ATM. The Internet's evolution from a technology for ensured[1] message delivery to a system offering QoS commitments needs to be questioned, especially if it should be able to compete with connection oriented systems. The questioning targets at the:

Ability to provide QoS in the Internet. According to thoughts in Odlyzko's *Paris Metro Pricing for the Internet* (PMP) [2], the concept of guaranteed QoS is a mirage, since iron-clad guarantees can only be given for bandwidths.
The usage and consumption of a shared Internet resource interferes always with the parties using the remaining resource portions. Thus commitments on QoS parameters like jitter and delays cannot be set in a allying way, since the interferences due to resource usage are unpredictable.

Necessity for QoS commitments. A plenitude of arguments for making the Internet QoS enabled, target at applications, which require a lot of bandwidth and which do only allow few constraints on time related QoS parameters, e.g. jitter, delay. Indeed few of these applications exist or are in frequent use nowadays. However, experience and history show that once the technology is mature to offer a platform, the applications will follow. The rising question isn't now if the Internet is able to offer this platform but what kind of platform it currently offers and what needs to be improved.
Actual investigation has polarized on the topics of congestion, i.e. the avoidance and control of congestion, and on QoS provisioning. Alas the only denominator of these two topics seams to be the requirement for charging and accounting. Both agree that future applications *will* increase traffic, but they differ in the decisions to manage the traffic. The QoS approach tries to satisfy the needs of the applications, whereas the congestion approach combats the occurring symptoms in the network. There is to admit, that congestion control and reduction influences

[1] Ensured in the sense that a large amount of devices can break down and delivery is still fulfilled.

positively the QoS offered to applications, but this is only a side-effect. The Achilles' heel of the congestion approach lays in the insufficient cognition of even nowadays congestion characteristics, i.e. one hardly finds information and statistics about real occurred congestion in the Internet. So design decisions for present and future Internet technologies heavily base on assumptions, which can artificially be provoked, but which do not necessarily correspond to the reality.

4.1 Congestion Management vs QoS Provision

As already mentioned the QoS approach targets at the requirements of applications of end–systems. They are expressed in QoS parameters and based on them resources are allocated throughout the network. The QoS approach is thus mainly concerned with signaling, admission control and QoS enforcement. It is a pure technical solution, implying that the customer primarily selects technical parameters, i.e. that she selects or specifies her application requirements in terms suited for her and that these terms will appropriately be mapped unto technical parameters, e.g. a layman will preferably use terms like *normal, good and excellent* to express quality, whereas a professional terms specifying *maximal jitter, minimal throughput etc.*. However, pricing needs to be introduced to moderate the customer in her delight. Resuming it means that the QoS approach primarily focuses on the needs of the *individuals*, i.e. individual user, individual application, but that for fairness reasons pricing needs to be introduced.
The congestion management approach applies on the network as a whole. The network is modeled as a limited resource accessed by a set of individuals, but which are often assumed to be a 'greedy' mass (or set of very delighted users). The goal is now to avoid and manage congestion by shaping and conditioning the traffic. Note that pure congestion control approaches neglect and avoid the usage of admission control (cf. PMP [2], ECN [13]). As a consequence and major difference to the QoS approach, the whole bunch of individuals, which provoke congestion, needs to be moderated. Some current protocols in frequent use like TCP, already own mechanisms for congestion management, indeed the approach absolutely needs to extended, since this kind of congestion management dissipates but does not avoid the occurrence of congestion. The extension is performed by introducing pricing mechanisms, which affect customer's behavior and thus actual network state. Thus congestion management solutions are mainly economical, what implies that they will be driven by market forces. Consequently engineering cannot give warantees that it will ever work.
Resuming, the congestion management approach primarily focuses on the network and the *mass* using it. The foundation is pricing, the approach is

fairer[2] than the QoS approach, since there is no admission control. Still the question remains if applications with few constraints on QoS will be applicable. As already said, little is known about today's congestion situation and less about the requirements of future applications.

5 Differentiated Services

As seen the characteristics of congestion in nowadays Internet are hardly analyzed, so the deliverable will focus on the QoS approach and on DiffServ as an example of the QoS approach.

DiffServ has primary been designed to push the complexity of reservations to the edges of the network. The scalability and thus efficiency of DiffServ founds on the discretization of the QoS[3] into some few classes, called service classes[4]. An application only obtains QoS commitments matching one of these service classes.The exception is the bandwidth, which has to be negotiated at communication begin.

In contrast to PMP, which is an example for the congestion management approach, and which founds on principles of a network divided in several technically equivalent but different in price channels, the division into classes (DiffServ)/channels(PMP) results in predictable improvements on QoS. Furthermore the interferences (cf. section 4 on page 5) on resource usage are limited due to DiffServ's provision of admission control. Therefore additionally to hard bandwidth warranties, soft guarantees on time-critical parameters can be made.

The requirements of Internet VPNs, especially of LAN–LAN VPNs, can sufficiently be satisfied by DiffServ. DiffServ is still a matter of investigation and several proposals on its implementation exist. To keep the setup of VPNs simple a DiffServ with the three service classes: Premium, Assured and best–effort is desirable. The characteristics of the premium service correspond to the ones of a leased line, it is thus also called a virtual leased line. Its drawback is the higher costs and the inflexibility to support bursts. The assured service corresponds to a service that has the same delays as the best–effort service, but that has a high probability to offer the negotiated bandwidth. Compared to the premium service it can handle bursts. Finally, the best–effort service is equivalent to the current Internet.

[2] Fair depends from the point of view, e.g. if all customers have equal rights to access the resource one can call it fair, whereas if a customer consuming a small amount of resources suddenly suffer losses of quality because another user is wasting unnecessarily immense amounts of resources, it is not categorically fair.

[3] The bandwidth excluded.

[4] In DiffServ a service or service class is defined by association of behavior aggregates and per hop behaviors, cf [10].

6 Charging and Pricing Differentiated Services

It is unavoidable to charge customers for the usage of differentiated services, because better communication quality can only be offered at the expense of the existing ordinary communication[5], e.g. suppose all customers would fairly and simoultaneously choose the best 'available' quality, they would get the ordinary best-effort one (in the case there is no admission control in this system). Charging forces customers to reflect on their needs. Charging is fair since customers getting better services, on expense of others, are paying for it. This fairness could imply a highly dynamic pricing model. Besides for fairness and regulatory reasons, charging revenues allow ISPs to amortize, maintain and extend their networks.
In the following subsections different pricing models will be examined and checked for their suitability for DiffServ–based VPNs. Since many different pricing models exist a classification scheme is introduced as well.

6.1 Classification Model for Pricing and Charging Systems

Pricing models have normally been developed to fit to some specific QoS providing system. Thus, different and usually incompatible pricing models exist, reflecting only rather modest standardization efforts. However, from the pool of existing pricing models basic classification and distinction may be deduced. Therefore, a classification based on five attributes and a number of well-defined parameter will be presented. Amongst other possibilities, pricing models can be classified according to the attributes *time, space, quality (i.e. class quality characterization), technological requirements*, and *volume* where each of these attributes shows a set of different parameters as summarized in figure 1.

Time	**Space Characterization**	**Class Quality Requirements**	**Technological**	**Volume**
duration	distance	ISP	flow–based	linear cumulation
period	route/path	customer	class specified	non–linear cumulation
time of day	location	self adjusting	not applicable	not applicable
not applicable	not applicable	indifferent		
		not applicable		

Table 1. Classification Model: Attributes and Parameters.

These parameters are discussed as follow. For the attribute *time*, parameter semantics have been applied:

- *Duration:* defines the elapsed time between start and end of service usage, e.g., duration of a video conference.

[5] In the case of Pareto efficiency

- *Period:* determines the committed length measured in time, which is per se independent, i.e. decoupled, of the service appliance. This commitment is usually set up in advance, e.g., a leasing period.
- *Time of Day:* defines the sensitivity of service usage to a given time of day. The influence of the time of day may be known in advance by the customer, e.g., weekend tariffs, or they as well may dynamically change, e.g., based on congestion in an auction system.
- *Not applicable:* means that the attribute *Time, Space, Class Quality Characterization, Technological Requirement*, or *Volume* is not relevant, e.g., for the time attribute not applicable means, that the time has no significance at all, which can be reasonable, e.g., in a volume–based system.

The attribute *space* allows for the following parameters:

- *Distance:* is the length of the (virtual circuit) from the sender to the receiver, which is passed by messages. Its length in meters is not relevant, but rather how much infrastructure has been used to enforce service provision.
- *Route/Path:* in contrast to distance the attribute route/path describes the relevance of where the message flow passes through, i.e. through which, how many, and what kind of routers. The route/path attribute plays in important role, when particular associations are made between the chosen circuit and the service.
- *Location: distance* and *route/path* parameters are not sufficient alone to describe all occurring cases for pricing models. Suppose edge-pricing (cf. subsection 6.2) has to be expressed. Saying that the distance and the route/path are not relevant implies a transparent network from the point of view of space. Indeed a transparent network (cloud) does not imply local importance of service provisioning. Thus location allows to consider places/entities in the network, which have particular importance for the pricing model, e.g., just like edge pricing.

The *class quality characterization* attribute describes the sensitiveness of pricing models for quality classes. It mainly explores how a differentiation of quality is made and who is taking influence on the selection or creation of quality classes, i.e. if only the ISP, the customer, or both. It has to be noted that a differentiation of quality does not imply that only a fix number of classes exists.

- *ISP:* sets up quality classes. Often the ISP will have a limited set of quality classes, which it may slightly adapt and distribute among customers.

- *Customer:* initiates and defines quality class specification, e.g., with a signaling protocol such as RSVP.
- *Self adjusting:* the class quality may change with network state, where the correction towards the new stable state is performed in a system inherent manner, e.g., as with PMP [2].
- *Indifferent:* no different quality classes available.

For the attribute *technological requirements* the following parameters exist:

- *Flow-based:* the supporting network technology offers a clear technology for maintaining flows within the network, such as with the Integrated Services Architecture
- *Class-based:* the network supplies a set of discrete classes, where the classes are not necessarily associated to particular technologies or QoS commitments.

Finally, the *volume* attribute distinguishes:

- *Linear cumulation:* the amount of data is accumulated linearly, determining that every single data unit measured has the same weight.
- *Non-linear cumulation:* covers all other cases, where the volume is taken into account of a pricing model.

It is obvious that the combination of all parameters gives raise to a large amount of different pricing models. It is up to the designers to agree upon the most reasonable ones. As already seen by the parameter *not applicable*, pricing models do not require to precise themselves on all attributes. In case that a pricing model has to chose just one parameter per attribute it is inappropriate. Therefore a supplementary notation is introduced. The following two variables x and X describe following alternatives of the importance of a given parameter per attribute:

- *x:* exactly one, but an arbitrary parameter of an attribute needs to be set, e.g., for the attribute *Volume*: x = [*linear cumulation* | *nonlinear cumulation* | *not applicable*]
- *X:* at least one, but an arbitrary number of parmeter(s) of the attribute needs to be set. This is required, if a combination of parameters is utilized to precisely define the scope of the pricing model, e.g., for the attribute Space: X= distance & route/path.

6.2 Edge Pricing

The design goal of DiffServ was to push complexity at the edges of the network, consequently the introduction of a charging system should not undermine this efforts. Since a charging system induces a considerable amount of complexity and has some forfeits on performance it is favorable to place it at the edges of the network.
This so called edge pricing system specifies only where to place the charging entities. It makes no assumptions on pricing policies and is therefore very flexible. In the classification model for pricing and charging systems, pricing models basing on edge pricing can qualify this property by setting the parameter *location* of the *space* attribute.

6.3 Volume–based Pricing Schemes

Volume–based charging applies prices to the amount of data transmitted. The scheme requires suitable metering components to monitor the amount of transmitted data. The concept has been used commercially to X.25 networks as well as the different service classes of ATM traffic during the 1996 tariffs in Switzerland. Most of these approaches use a system of price discounts based on several thresholds. For Internet traffic applying a volume–based scheme, the two examples of the traffic metering approach in New Zealand [14] and Great Britain [15] are already well-known. Moreover there are approaches to model accurately the relationship between the current utilization of a resource and the price to be paid for using it (cf. [16]). A static volume–based model with edge pricing may look like:

Time	**Space Characterization**	**Class Quality Requirements**	**Technological**	**Volume**
duration	location	X	not applicable	x

6.4 Auction–based Pricing Schemes

A congestion pricing scheme (where packets are charged if and only if the network is congested) can be implemented by using a "smart market", where the price for sending a packet varies on a very short time-scale, thus reflecting the current degree of network congestion. Each packet header contains a bid field, and the packet is admitted, if the bid exceeds the current marginal cost of transportation.
An example for auction systems is the Generalized Vickrey Auction. There the highest bidders win, but pay only the market-clearing price. The Vickrey auction approach gives all competing customers an incentive to disclose their true evaluation of the good [17]. The mechanism guarantees only relative priority, no absolute QoS. It has to be noted that the smart market model as well as auctions do show an often quoted drawback in terms of lacking price transparency and predictability, which results, e.g., in problems for communication budget definitions. Improvements of the Generalized Vickrey Auction can be found in Delta Auctions [18] and the CHiPS approach [19]. A specific Auction–based pricing model could look like:

Time	Space Characterization	Class Quality Requirements	Technological	Volume
time of day	route/path	self adjusting	flow–based	non linear cumulation

6.5 Paris Metro Pricing (PMP)

The Paris Metro Pricing (PMP) model is an excellent example for the *congestion management approach* described in section 4 on side 5. PMP suggests to satisfy the requirements of differentiated services by splitting networks in logically separate channels, where the traffic over these channels is all treated as best–effort (in the case of the Internet). The core idea presumes that the channels differ in their price and due to this, more expensive channels will have less traffic and will thus provide better services. An example for a PMP specification could be:

Time	Space Characterization	Class Quality Requirements	Technological	Volume
duration	x	indifferent	class–based	x

6.6 Dynamic Pricing Schemes

At the beginning of the section 6 some thoughts on pricing and fairness have been sketched. A postulation was made where fairness and customer interference could imply the usage of dynamic prices. To clarify the idea: the competition of customers for a limited resource reflects itself in a raising or falling worth of the resource in use, according to the number of customers in competition. Dynamic pricing can be very similar to auction pricing schemes, but with the difference that prices are not set by customers but by the ISP and possibly only under certain conditions, e.g. application of prices in a router as soon as congestion occurs (ECN). Thus dynamic pricing as well as the other just presented pricing schemes represent only an generic term to accentuate special characteristics. A dynamic pricing scheme could be classified as:

Time	Space Characterization	Class Quality Requirements	Technological	Volume
time of day	location	ISP	not applicable	x

6.7 Flat Rate Pricing Scheme

In a flat rate pricing scheme the customer pays a fix price for the potential use of the network infrastructure over a fix time period. The price the customer is paying for, is absolutely independent from duration, amount and how often the infrastructure has been used. A typical flat rate can be classified as follows:

Time	Space Characterization	Class Quality Requirements	Technological	Volume
period	not applicable	ISP	x	not applicable

According to experience and several inquiries, flat rates are preferred by consumers. Furthermore, flat rate pricing often allows ISPs to collect more

revenue (cf. [2]), since consumers are willing to pay more for predictable pricing. Additionally a pricing system for flat rate does not need metering and charging infrastructure[6].
However, flat rate systems have the disadvantage not to provide economic congestion control mechanisms e.g. comparable to PMP, for resource allocation resulting in that flat rate is not efficient. Furthermore differentiated services stringently require usage sensitive charging as a fund for resource protection and customer incentive to select the differentiated service suiting the application requirements. Resource protection means thus, that the resource is protected from customers that are not *explicitly* willing to pay for it, i.e. customer would not use the resource if he had to pay for it. Flat rate undermines these efforts because customers take long–time decisions e.g. select monthly of rate/service offer. Based on the taken decision he will in his delight, consume resources that his applications not necessarily require and that he would never use, if he had to pay for, e.g. running FTP as a highly priorized service, due to having payed a flat rate for high priority services besides to the best–effort services. Thence differentiated services require differentiated pricing schemes. Establishing different flat rates for each service of a differentiated service architecture will not suffice, since this will only provide small resource protection.

7 Balance on Pricing Schemes

Usually flat rate pricing is associated as a pricing scheme offering customers predictable costs independently of usage. Despite higher revenues to ISPs networks are not optimally used, since customers pay for more as they use and use more as they should, at least for most applications. This kind of flat rate pricing has mainly no regulator effects in the sense of resource consumption and customer behavior, due to its static characteristics.
Dynamic and volume–based pricing schemes[7] are the antagonists of flat rate pricing. Dynamic pricing schemes originate from a global view of actual network characteristics with the accentuation to *TIME(strong) ,volume(weak)* parameters. On a supply and demand foundation prices are set and customers chastened. Because all users are affected, dynamic pricing schemes are said to be fair, but that implies not that this is desirable. Each single action of a customer affects automatically all other customers actually using the same network resources. Dynamic pricing schemes supplementary couple customers to each other. From the point of view of a customer and its applications prices become unpredictable and from the engineering point of view

[6]Remark: from the telephony systems there is said, that the overhead of the account infrastructure consumes approximately 50% of the phone bill.
[7]Auction–based pricing schemes, as well as PMP will not be examined anymore in these investigation.

coupling is seldom desired since it complicates the analyze of systems (cf. *divide et impera* philosophy). Additionally dynamic pricing schemes, since prices are not predictable, do not fit well in the policy to shift traffic to time periods where the network is hardly used, i.e. shift traffic like updates and mirroring to the night. Besides dynamic pricing is unlikely complicated to implement.
Volume–based (accentuation: *time(weak), VOLUME(strong)*) pricing schemes focus mainly on the usage of resources over a time period, so their regulator behavior is rather static, i.e. not sensitive on short-term resource shortage. The chastening of customers is done by accounting the amount of data sent, so prices are roughly predictable. The scheme is less fair than dynamic pricing, since customers are decoupled from each other, except from interferences of resource usage. Shifting of traffic to less frequent used time periods can easily be done by introducing different pricing classes, e.g. as done in the telephony by distinguishing tariffs for day and night. From the implementational point of view, volume based pricing schemes are less complicated to implement than dynamic pricing. Nevertheless they introduce several new entities like meters, accounting and charging units in the network, resulting in higher management and maintainance efforts for ISPs.

Concluding there can be said:

- flat rate pricing schemes offer predictable prices and are easy to implement.
- flat rate pricing schemes lack of chastening effects.
- dynamic pricing schemes are fair but too complicated.
- dynamic pricing schemes affect the whole network and couple customers behavior.
- volume–based pricing schemes are less fair than dynamic pricing ones.
- volume–based pricing schemes decouple customers behavior.
- volume–based pricing schemes still are complicated.

The drawback of nowadays commonly used flat rate pricing compared to dynamic and volume–based pricing consists in its static nature and thus its lack of adaptive regulation. The advantage in its predictability, thus high customer acceptance and its simplicity.

8 Pricing Scheme for DiffServ–based VPNs

Flat rate pricing are what customers want, but flat rate undermines the efficient usage of DiffServ. To overcome the drawbacks and to savor the

advantages of the flat rate pricing scheme a dynamic and adaptive flat rate pricing is developed.
Basically one can say, the conception of dynamic and volume–based pricing founds on theirs combination of time and resource consumption, i.e. dynamic pricing on the *current* resource consumption and volume based pricing on the resource consumption over a time period. By now flat rate lacks of this time and consumption catenation, since it only uses a time component, namely the time period where the flat rate is valid.
The idea consists in additonaly making flat rate pricing volume sensitive without losing the advantage of price predictability and by keeping a simple solution in mind. A periodical adaption of the flat rate pricing according to the volume passed during several billing periods will already fulfill the quest.

$$price_0 = \text{ISP\_Offer} \quad (1)$$

$$price_i = price_{i-1} + \begin{cases} \Delta_{price} & \text{if } i \bmod n = 0 \\ 0 & \text{else} \end{cases} \qquad i, n \in \mathbf{N} \quad (2)$$

At a customer request for VPN setup, the ISP and the customer enter into a negotiation phase. The negotiation targets at seizing the probable characteristics of the traffic the customer will generate. Based on these assumptions the ISP makes an flat price offer (cf. equation 1). After several accounting periods, noted by i in equation 2, the flat rate price is adapted (Δ_{price}) according to the effective traffic characteristics of the customer. The adaption period is n times longer than the accounting periods. In the case of $n = 1$ flat rate pricing is most adaptive, i.e. the adaption period is equal to the billing period. This special case almost corresponds to the volume–based pricing scheme with an additional delay of one billing period. In the case of $n = \infty$ the adaptive component, i.e. the consumption indicator, is lost. This special situation corresponds to the actual non–adaptive flat rate pricing scheme.
The idea of an adaptive flat rate pricing scheme is not new. It is widely applied for pricing gas and electricity consumption, e.g. for gas consumption the monthly bill, which is a flat rate, is refined on the consumption of the previous year, i.e. i = month, n = 12 (year). Experiences in these branches show, that adaptive flat rate pricing, due to its sensitiveness for consumption volume, motivates customer chastening by simoultaneous price prediction. The scheme is furthermore easily understood and thus accepted by consumers. Additionally it is technically easy, compared to dynamic pricing and volume–based pricing, to implement since there is only need for metering units at the edge of the network, i.e. at the customer's access points, e.g. in the gas example the meter only needs to be consulted once a year to adapt and fix the prices for the coming year.
From the point of view of DiffServ this bonus malus system is a differentiated

pricing flattened over a long period, but which nevertheless optimally affects customers behavior.

9 Service Level Agreement (SLA)

The set up of temporary networks starts with the estimation of the characteristics of the traffic, which will be transmitted, the estimation of the QoS required and possibly the living time of the temporary network. Assuming the temporary network is built on DiffServ technology, the estimation phase is equivalent to a negotiation phase, where customer and ISP probe each others needs and facilities with the goal to find a consensus. The consensus is expressed in a Service Level Agreement (SLA), which is a contract consisting of legal, financial and technical commitments. Recommendations on terminology and protocol definitions to set up SLAs are hardly defined. However unisonous opinion exist, that SLA negotiation will not be automated but involve always human interaction.
From the SLA, which includes commitments but not technical details, a Service Level Specification (SLS) is deduced. The SLS contains the technical details for service definition and provisioning. SLSs have the advantage that they can be negotiated and set up automatically, i.e. between the communication parties without any user intervention, what is especially advantageous for dial–up and mobile customers. A part of the SLS includes also parameters required for the respective pricing model. In the case of the adaptive pricing model presented in the previous section the parameters are for example the duration of a billing period and the refinement interval required for the price adaption.

10 Conclusion

DiffServ is an appropriate foundation for building Internet VPNs, since the characteristics of a leased line can approximately be recreated. The advantages are the intense cost reduction and the ability to tailor services according to the requirements of companies.
The necessity for charging and accounting VPNs is satisfied by using an adaptive flat rate scheme, which offers flexibility and acceptance of customers by appliance of price prediction and acceptance of ISP by price adaption in case the contract, i.e. the SLA, changes.
Future investigation will focus on further analysis of adaptive flat rate pricing and its integration in DiffServ.

References

[1] J. Harms: *Advanced Network and Agent Infrastructure for the Support of Federations of Workflow Trading Systems.*
Froschungsgesuch Abt.IV, NFP/SPP.

[2] A. Odlyzko: *Paris Metro Pricing for the Internet.*
May 25, 1999.

[3] K. Nichols, V. Jacobson, L. Zhang: *A Two-bit Differentiated Services Architecture for the Internet.*
April, 1999.

[4] F. Baumgartner, T. Braun, P. Habegger: *Differentiated Services: A New Approach for Quality of Service in the Internet.*

[5] J. Lindström: *How will accounting and billing work in the next generation of the Internet?*
17.4.1999

[6] L. W. McKnight, J. P. Bailey: *An Introduction to Internet Economics.*
March, 1995

[7] H. R. Varian: *Differential Pricing and Efficiency*

[8] *QBone Bandwidth Broker Architecture.*
http://qbone.ctit.utwente.nl/BBroker/bboutline2.html

[9] K. Nichols, S. Blake, F. Baker, D. Black: *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers.* RFC 2474, December 1998.

[10] S. Blake, D. Black, M. Carlson, E. Davis, Z. Wang, W.Weiss: *An Architecture for Differentiated Services.* RFC 2475, December 1998.

[11] J. Heinanen, F. Baker, W. Weiss, J. Wroclawski: *Assured Forwarding PHP Group.* RFC 2597, June 1999.

[12] V. Jacobson, K. Nichols, K. Poudri: *An Expedited Forwarding PHP.* RFC 2598, June 1999.

[13] K. Ramakrishnan, S. Floyd: *A Proposal to add Explicit Congestion Notification (ECN) to IP.* RFC 2481, January 1999.

[14] N. Brownlee: *Internet Pricing in Practice.* In: L. McKnight, J. Bailey (edts.) Internet Economics; MIT Press, Cambridge, Massachusetts, U.S.A, 1997, pp 77-90.

[15] R. Rogerson: *Usage-related Charges for the JANET Network.* JISC Circular 3/98, March 1998, available at the URL: *http://www.jisc.ac.uk/pub98/c3_98.html*

[16] P. Reichl: *Kelly's Bound, RUAA and the Pricing of Multiclass Traffic in Loss Networks.* U.K. Performance Engineering Workshop, Bristol, U.K., July 1999.

[17] MacKie-Mason, H. Varian: *Pricing the Internet.* In Public Access to the Internet, B. Kahn, J. Keller (edts.), Prentice Hall, Englewood Cliffs, New Jersey, U.S.A.,1995.

[18] G. Fankhauser, B. Stiller, C. Vögtli, B. Plattner: *Reservation-based Charging in an Integrated Services Network.* 4th INFORMS Telecommunications Conference, Bocca Raton, Florida, U.S.A., March 1998.

[19] P. Reichl, G. Fankhauser, B. Stiller: *Auction Models for Multiprovider Internet Connections*; Messung, Modellierung und Bewertung von Rechensystemen (MMB'99), Trier, Germany, September 21-22, 1999.