Provider-Oriented Linear Price Calculation for Integrated Services

Martin Karsten[1], Jens Schmitt[1], Lars Wolf[1], and Ralf Steinmetz[1,2]
[1] Darmstadt University of Technology *
[2] German National Research Center for Information Technology, GMD IPSI
Email: {Martin.Karsten,Jens.Schmitt,Lars.Wolf,Ralf.Steinmetz}@KOM.tu-darmstadt.de

Abstract

When reservation-based service differentiation is offered in an IP network, it is crucial for a network provider to appropriately charge for service invocations. For internal calculation as well as external price representation, a notion of costs and prices for communication services is needed. In this paper, general requirements for cost and price calculation are analysed for packet-switched multi-service networks. We conclude that internal price calculation should be linear, based on resource usage and uniform across multiple service classes. Accordingly, we specify refinements for service class definitions and present a price calculation method for the IETF's Integrated Services architecture. The application of such a price representation to existing pricing and charging approaches is shown. Finally, certain economic aspects of the *Guaranteed* service class are analysed and the results are expressed using linear price representation.

Keywords Integrated Services, Charging, Costs, Prices, Calculation, Business Model, QoS

1 Introduction

The transition of the Internet towards a commercially funded and used integrated services network raises, among others, the question about how network usage can be charged appropriately. Clearly, the current charging schemes (mainly flat-fee access-based or time/volume-based) will not be sufficient in the presence of multiple service classes, resource reservation and discrimination between different usage requests [MMV97]. From an economic point of view, communication services are characterized by:

- availability of a non-storable resource (network capacity)
- high fixed costs & low variable costs

In economic theory, these characteristics, which are similar to traditional telephony, electricity, aircraft seats, etc., are dealt with by using a management technique called *Yield Management* [Lei98]. When Yield Management is used, calculation is based on *profit contribution* and *opportunity costs*, instead of using full-cost or variable-cost calculation. The concept of profit contribution means that each resource unit is sold for a price higher than its marginal cost. The difference of both contributes to the overall revenue, which must exceed the overall investment for the appropriate business cycle. Opportunity costs describe the fact that selling a resource unit prohibits using it for another business transaction. Under given price-demand patterns, prices can be optimized to maximize the overall revenue.

The IETF's Integrated Services (IntServ) architecture [BCS94] defines multiple service classes for packet-switched network communication. For each class, quality of service (QoS) is described by a vector of partially different parameters. All service classes compete for the same underlying network resources, such that an internal calculation model should be related to resource usage. However, service requests cannot directly be compared with respect to resource consumption, especially across service classes. Therefore, the existence of multiple service classes presents new challenges to a cost and price calculation model, which are only partially addressed by existing work. In this paper, we attempt to formulate these challenges and specifically analyse the IntServ model and service classes with respect to the resulting requirements.

We begin by briefly reviewing existing related work in Section 2, then we give a short introduction to provider-oriented price calculation in Section 3. In Section 4, we extend the point of view to multiple service classes in a packet-switched network. Afterwards, we present our model for price calculation in Section 5 and show applications of this method in Section 6. In Section 7, an analysis of certain economic aspects of the *Guaranteed* service class is given. Finally, we summarize our findings and give an outlook on further research issues in Section 8.

---

*. This work is sponsored in part by: Volkswagen-Stiftung, Hannover, Germany and by Deutsche Telekom AG, Darmstadt, Germany.

2 Related Work

So far, significant research work has been published on the issue of pricing in telephone networks (see e.g. [MV91] and references contained herein). In the area of packet-switched multi-service networks, approaches to find welfare-optimizing price models can be found in e.g. [MM97,PSC98,SFY95]. On the other hand, little work has been done to find a provider-oriented calculation model for packet-switched multi-service networks. A very general, complete and hence complex price calculation model has been presented in [WPS97], although it lacks full applicability to multiple service classes. Particularly, details about the relation of prices to resource-oriented admission control are not considered in [WPS97].

It has been established by many researchers [MMV97,WPS97,SFY95] that pricing based on real marginal cost is not sufficient for communication networks. From an economic point of view, congestion costs are often used as marginal costs, i.e., consumption of resources prohibits other users to use them and lowers their utility. The goal of a pricing algorithm is then to provide optimal resource allocation with respect to the users’ utility function [MMV95, SFY95, GSW95, KVA98].

In this paper, the problem of appropriately pricing integrated services networks is examined from the opposite point of view. This, we believe, is a highly realistic approach, since the network provider’s context might eventually determine price strategies. From a business management perspective, optimal welfare is not the primary goal of pricing. Instead of users’ utility functions, a network provider only experiences their willingness-to-pay. In that sense, marginal costs have to be combined with opportunity costs and optimization of profit becomes the overall goal [MV91,WPS97]. Our work is based on previous results from [KSWS99].

3 Single-Service Provider-Oriented Price Calculation

Certain restrictions are usually applied to a calculation model to keep the economic time horizon limited and the complexity of the overall problem tractable. Specifically, we employ the notion of a *business cycle* as a time period in which a specific investment volume has to be recovered. We further assume the existence of an *aggregated* price-demand estimation for each time during the business cycle. A simple profit-maximizing calculation model for a network providing a single communication service class can be expressed as follows [MV91]:

$$\int_0^{Tb} R(\gamma(t))dt - K(C) \qquad (1)$$

under the constraint:

$$\gamma(t) \le C \qquad t \in [0,Tb] \qquad (2)$$

Variables used:

$\gamma(t)$	aggregated demand at time t
$R(\gamma(t))$	aggregated revenue at time t
Tb	duration of business cycle
C	total available resource capacity
K(C)	amortization of capital investment over one business cycle

In such a model, the time parameter is a constant scaling factor, i.e., price and demand are applicable per fixed time unit, which can be chosen arbitrarily small. In the big view, a calculation model can be used in two areas of the cyclic calculation and planning process, shown in Figure 1. First, during capacity planning, network capacity can be increased as long as each increase is covered by expected revenue for this investment. However, changing a communication network’s capacity happens on a rather long time-scale, therefore a second application is to use this calculation model to optimize revenue under a given limited capacity. In the second case, opportunity costs come into play when there is more demand than supply. If a service request has to be refused because another service request occupies resources, then the potential revenue of the refused request can be considered as opportunity costs of the accepted one. Although opportunity costs are not directly expressed in (1), they are implicitly included when optimizing it. In general, during capacity planning as well as network operation, a certain target value can be expected to limit the fraction of resources usually available for reservations. This is desirable, for example, to keep the blocking probability low or to prohibit starvation of unreserved traffic.

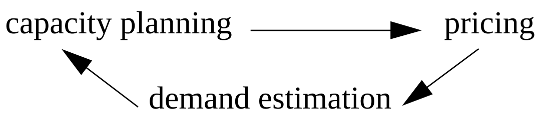

Fig. 1: Cyclic dependency among calculation tasks

4 Multi-Service Provider-Oriented Price Calculation

There are several constraining factors for price calculation of packet-switched multi-service networks, resulting from multiple service classes using the same resources. In the following subsections, we first establish two axiomatic requirements, then we present general characteristics of a calculation model resulting from these requirements. Finally, we list additional aspects, which strongly devise cost and price calculation to be modelled according to the axiomatic constraints.

Under a profit-contribution model for communication services, the terms *cost* and *price* can be used somewhat interchangeably from the network provider’s perspective. Marginal costs are negligibly low and in case of limited capacity, opportunity costs basically equal prices and are implicitly included when such a model is optimized. We assume that actual market prices consist of a transaction component, a resource reservation component and possibly other components. Furthermore, actual prices can be influenced by marketing considerations and deviate from calculated prices. Nevertheless, it is important to internally use precise calculation as a reference model for the daily business process. We use the term *price calculation* to refer to internal cost and price calculation for the resource reservation component.

4.1 Axiomatic Constraints

Linearity The price component for resource usage must be linear.

$$a \cdot p(x) = p(a \cdot x) \qquad \text{for resource usage } x \qquad (3)$$

This requirement is due to the possibility of arbitrage (resale at low transaction costs) in packet-switched communication networks. Because the service of transmitting packets from one node to another can be used for many different applications, it might be hardly feasible for regulation authorities to prohibit arbitrage in a conventional (i.e. legislative) way. Furthermore, the objective of prohibiting resale might be dropped. Any non-linear pricing scheme, however, can be exploited by arbitrage [MV91]. Even if external arbitrage is not an issue, linear resource prices seem to be most appropriate for internal calculation, because they properly reflect resource consumption.

Uniformity The price component for each resource’s usage must be uniform across service classes.

$$p_{S1}(x_q) = p_{S2}(x_q) \qquad \text{for each resource } q \text{, usage } x$$
$$\text{for each pair of service classes } S1,S2 \qquad (4)$$

There are two reasons for this axiom. First, requests for different service classes might be substituted by customers, exploiting the knowledge about a service class’ definition. This can be done for immediate use or for resale. Second, opportunity costs might apply across service classes, if a request for one service class has to be refused, because a request for another service class occupies the resources.

4.2 General Characteristics

A common requirement for pricing communication services is that prices should be known before a service is requested [FD98,FSVP98,KSWS98], hence the price per resource unit must be stable during a specific time period. Assuming this and taking into account both axioms of Section 4.1, an internal price calculation for multi-service networks should be defined as follows:

$$p(x,t) = a_1x_1 + a_2x_2 + \ldots + a_nx_n \qquad \text{for complete}$$
$$\text{resource vector } x = (x_1,x_2,\ldots,x_n) \text{ at time } t \qquad (5)$$

This linear price formula must be used for all service classes. For capacity planning it can be used to determine the optimal network capacity. In case of limited capacity, the optimal traffic mix during peak-load periods can be estimated similarly. This calculation method is called *linear price calculation*.

4.3 Further Considerations

Auctions From an economic point of view, every sales transaction can be considered as an auction [MM98]. Winner determination takes place by ordering all bids and choosing the highest one(s). In case of a multi-service network, multiple resources have to be considered, which is called *combinatorial auction*. The underlying theoretical problem of winner determination in combinatorial auctions is proven to be NP-complete [RPH98], but approximate solutions exist [San98]. The problem becomes even harder, if bids from multiple service classes for multiple resources cannot be ordered at all. However, if cost and price calculation is uniform for all service classes, this additional complexity is resolved.

Demand Interdependence For packet-switched multi-service networks, it is very likely that demand patterns are interdependent between different service classes, because users might combine or substitute traffic flows of multiple service classes within a single application. Additionally, interdependency between resource parameters can exist. As an example, for a delay guaranteed service, the amount of buffer needed is largely determined by bandwidth and delay characteristics. If the price for a critical resource, e.g. bandwidth, is increased, demand for the service class, and thus demand for other resources, decreases as well. If a uniform price function is used, representing all resource parameters, per-flow demand estimation for each service class can be eliminated and replaced by estimating aggregated demand interdependency per resource parameter. This seems to be easier to accomplish, based on past measurements and experiences.

Multicast A thorough study of allocating costs among members of a multicast group is presented in [HSE97]. Cost allocation is described by splitting each link’s costs among a defined subset of group members. Definition of the subset determines the allocation strategy. Of course, the sum of all cost fractions must equal the total costs for a link. Realizing such an approach becomes much simpler, if costs can be expressed as a linear function of resource parameters, especially if charges are shared among receivers with heterogeneous QoS requirements.

5 Linear Calculation for Integrated Services

The most evident obstacle for developing a calculation model for the IntServ architecture is that service classes are currently not precisely defined. Therefore, we begin by refining the service definitions. The other details of each service definition can be found in [Wro97b], [SPG97] and [GGPR96], respectively. For reasons of brevity and simplicity, we do not explicitly consider the effects of traffic distortion for this model, other than what is specified in the error terms during Guaranteed service negotiation. Initially, we focus on a single link or a specific path through a network cloud, connecting two IntServ-enabled routers, although this restriction is partially dropped in Section 7. We briefly list the relevant flow specification and error term parameters [Wro97a] which are used throughout this paper:

b	bucket depth
p	peak rate
r	token rate
M	flow MTU
R	service rate
S	slack term
C	rate-dependent error term
D	rate-independent error term

The approach described in this paper presents a general idea for a calculation method, rather than specific calculation rules to be used for each implementation. However, we believe the service definition refinements given in the next subsection to closely resemble realistic services.

5.1 IntServ Service Classes

Guaranteed For the duration of a service invocation, each router is guaranteed to always have service rate R available for a flow conforming to the requested token bucket. Furthermore, all incoming packets exceeding rate R are forwarded within the (indirectly) specified delay bound and no conforming packets are dropped due to buffer overflow.

Guaranteed Rate A flow is guaranteed to be serviced with average rate r. No buffering is guaranteed.

Controlled Load A flow conforming to a token bucket is forwarded almost without queuing delay or loss, as long as its data rate is not higher than r. Bursts are forwarded with as little queuing delay and loss as possible, depending on the actual load situation. This is achieved by allocating a specific excess service rate and buffer for each flow and enabling flows to borrow unused resources from each other. An implementation might, for example, use a *guaranteed rate scheduler* [GLV95] in conjunction with *hierarchical link sharing* [FJ95] to accomplish such forwarding. We assume constant excess parameters f and g to be used for each flow (depending for example on the total number of flows and desired failing probability) and define the following simple formulas for the amount of service rate R and total buffer B:

$$R = r + (p - r) \cdot f\,, \quad B = b \cdot g$$
$$\text{with } f, g \in [0,1] \tag{6}$$

Please note that the following considerations do not rely on exactly the above definitions and formulas, but only on having any precise specification of resource usage in the first place.

5.2 Virtual Rate Parameters

In reality, only one parameter (service rate, i.e., forwarding capacity) denotes the total available service rate of an outgoing link. However, there are up to two rate parameters, r and R, in IntServ service specifications with even different semantics depending on the actual service class. In order to allocate costs to reservation requests, we therefore establish a resource model using three *virtual rate parameters*:

- The *token rate* (q_T) describes the forwarding rate that is always available and expected to be constantly used by a flow.
- The *clearing rate* (q_C) denotes a guaranteed forwarding rate on top of the token rate that is reserved per delay-guaranteed flow, but expected to be used only for bursts of data.
- The *residual rate* (q_R) is a forwarding rate on top of the token rate, which is only statistically available to a flow. This resource can consume the unused capacity of q_C.

These parameters can be used to express the resource consumption of service requests by mapping the rate parameters r and R from an IntServ flow specification to the virtual rate parameters according to Table 1. This mapping follows directly from Section 5.1 and the definition of virtual rate parameters.

service class	q_T	q_C	q_R
Guaranteed	*r*	*R - r*	-
Controlled Load	*r*	-	*(p-r) · f*
Guaranteed Rate	-	-	*r*

Table 1: Rate allocation for IntServ service classes

5.3 IntServ Calculation Model

Considering buffer space as additional resource parameter, we can establish a linear function

$$p(x_T,x_C,x_R,x_B) = a_T x_T + a_C x_C + a_R x_R + a_B x_B \qquad (7)$$

to assign resource consumption respectively prices to a flow requesting token rate x_T, clearing rate x_C, residual rate x_R and buffer space x_B. Prices are applicable per fixed time unit, which can be chosen arbitrarily small. The formula for calculating the amount of buffer B for Guaranteed service is given in [SPG97]. Converted back to the original IntServ parameters by using the definitions in Section 5.1 and Section 5.2, the price function for each services class can be expressed as follows:

$$\begin{aligned} p_G(r,R) &= p(r,R\text{-}r,0,B) &&= a_T r + a_C \cdot (R\text{-}r) \\ & &&\quad + a_B \cdot B && (8) \\ p_{CL}(r) &= p(r,0,(p\text{-}r)f,bg) &&= a_T r + a_R \cdot (p\text{-}r) \cdot f \\ & &&\quad + a_B \cdot b \cdot g && (9) \\ p_{GR}(r) &= p(0,0,r,0) &&= a_R \cdot r && (10) \end{aligned}$$

These price functions form the basis for an IntServ calculation model, which is linear and uniform across multiple services classes and therefore, fulfils this important requirement derived in Section 4.2.

For certain scheduling approaches (see [Zha95]), *schedulability* is an additional internal resource parameter. However, the service classes currently under consideration for Integrated Services heavily rely on rate-based scheduling semantics. Particularly, for Guaranteed service, each scheduler has to approximate a rate-based scheduling behaviour. Therefore, we do not explicitly consider schedulability as separate resource.

6 Applications of Linear Calculation

Linear calculation can be used for internal price calculation as well as external price representation. In this section, examples for both applications are given.

6.1 Optimal Pricing

The authors of [WPS97] present a very general and complete model for optimal pricing of multiple guaranteed service classes under consideration of price-demand functions. It is correctly pointed out there that analytically solving the whole model is mathematically intractable, therefore an approximating procedure is described to carry out planning and calculation. While other research approaches often deal with optimal pricing in a sense of optimal welfare, this pricing scheme is targeted to maximize profit for the provider. However, as noted in [WPS97], a similar model can be developed to maximize other objectives. By slightly modifying and applying linear calculation for IntServ service classes, we simplify and, at the same time, enhance the model in several ways:

- We concentrate on reservation-based services, therefore best-effort traffic is not explicitly considered in our model.

- Instead of using very general assumptions about admission control and the properties of service classes, we specifically consider the definition of IntServ service classes by using resource-based parameters and linear calculation. Thereby, the applicability to multiple real service classes is given.
- In [WPS97], communication services and demand patterns are modelled by the notion of calls, i.e., call probability, call duration, static QoS, etc. While being applicable to ATM service classes, this model does not fit well with the IntServ framework. Instead, our model only uses aggregated demand functions for each time period, i.e., no assumption about specific flows is needed, but only an overall estimation of aggregated demand per network resource, depending on the price-vector. By that, the new model implicitly encompasses the above details and also covers dynamic QoS.
- As mentioned in the article, [WPS97] does not cover interdependency among service classes and furthermore implicitly assumes a discrete set of service classes. In our model, the fact that each service class offers a vector-space of resource quantities is taken into account, although the estimation of demand interdependency between resources remains as an open issue.

In accordance with Section 3, the price function from (7) is extended by a time parameter to express different price settings at different points in time. The core formula which shows the total revenue that is to be optimized can then be specified and looks as follows (roughly using the notation of [WPS97]):

$$\int_0^{Tb}\left\{\sum_{X = T, C, R, B} a_X(t)\cdot\gamma_X(t)\right\}dt - K(C) \tag{11}$$

under constraints

$$\gamma_T(a_T(t), t) \le C_{TCR} \qquad t \in [0,Tb] \tag{12}$$

$$\gamma_C(a_C(t), t) \le C_{TCR} - \gamma_T(a_T(t), t) \qquad t \in [0,Tb] \tag{13}$$

$$\gamma_R(a_R(t), t) \le C_{TCR} - \gamma_T(a_T(t), t) \qquad t \in [0,Tb] \tag{14}$$

$$\gamma_B(a_B(t), t) \le C_B \qquad t \in [0,Tb] \tag{15}$$

Variables used:

$a_X(t)$	price coefficient for each unit of q_X at time t, corresponding to (7)
$\gamma_X(t)$	aggregated demand for q_X at time t, for price vector (a_T, a_C, a_R, a_B)
Tb	duration of business cycle
C_{TCR}	total available service rate (reservable bandwidth)
C_B	total available buffer space
K(C)	amortization of capital investment over one business cycle

Constraints (12), (13) and (14) denote the fact that the amount of service rate reserved as token rate cannot be reused, whereas service rate used as clearing rate can be used simultaneously as residual rate. Constraint (15) states that buffer space cannot be reused. Please note that this is in no contradiction with multiple Controlled Load flows borrowing resources from each other.

Comparing (11) with the corresponding formula in [WPS97] shows that using virtual rate parameters and considering only aggregated demand significantly reduces the mathematical complexity but nevertheless enhances the level of detail by considering real resources instead of a general admission control expression. While being subject of ongoing work, it is our assumption that in such a way, the problem of optimal pricing might be analytically tractable. We are convinced that in general such an approach is very useful to apply theoretic results in a real environment.

6.2 Application to Charging Mechanisms

In [KSWS98], an approach to exchange charging information between RSVP routers is presented. The problem of appropriately representing prices was left open for further study. Using the linear cost functions as price representation, we can establish a concise notation for prices which fits with the protocol-oriented approach of [KSWS98]. Although in [KSWS98] it was assumed that price representation probably depends on the service class, we can now formulate a single price function representing all service classes considered in this paper:

`price :=` `price for` q_T
`price for` q_C
`price for` q_R
`price for` q_B
`other charge components`

Using this notation, all necessary QoS-dependent price information is transmitted. There might be other charge components, for example a flow setup fee. This is represented by the generic field <`other charge components`>. Prices can be accumulated at each hop and because the price function is linear, upstream charges can easily be split at multicast branches (see also Section 4.3).

Note that even when the charge coefficients for each router are largely stable, it is usually necessary to transmit price information with each PATH message (see [KSWS98] for details). According to the *Edge Pricing* paradigm [SCEH96], the price function expresses the total accumulated charges from the sender to the respective next hop. Therefore, accumulated price functions for different flows using different paths are very likely to differ.

It is clear that an indirect price representation like this adds additional complexity to the end systems, in that this price representation has to be translated into a user-friendly format. However, translation of QoS parameters has to take place for IntServ requests anyway and it is a common design paradigm in the Internet to push intelligence towards the end systems while letting the network technology be as simple as possible. Therefore, we do not believe this slight additional complexity to be a major problem.

7 Economic Aspects of Guaranteed Service

Linear calculation is introduced as an approach to model resource and price calculation for single IntServ routers and attached links. In this section, we show how this modelling technique can be exploited to further analyse certain economic aspects of the Guaranteed service class. This also demonstrates how linear calculation can be extended to eventually cover a whole network domain consisting of multiple IntServ routers and respective transmission lines, and to analyse economic end-to-end aspects.

For these examinations, we assume a charging model where charges from each router are accumulated and shared between sender(s) and receiver(s), similar to [KSWS98,FSVP98,CSZ98]. The values of price coefficients a_X used in this section are not assumed to be globally uniform, instead they are local to each router. The service rate R is assumed to always be larger than token rate r.

7.1 Token Rate vs. Clearing Rate

In [DVR98], it is pointed out that a receiver might choose to lower its service rate requirements R by increasing the average data rate *r* when requesting Guaranteed service. Since q_C, the difference of R and *r*, can be used for providing Guaranteed Rate service and Controlled Load service, a pricing scheme should give the right incentives for users to choose r according to their average data rate. Using linear calculation, this can be achieved by setting an appropriate higher price for q_T than q_C. The economically optimal price relation between q_T, q_C and q_R is part of the optimization problem from Section 6.1.

7.2 Error Terms

The C and D error terms, which are part of Guaranteed service negotiation, partially determine the service rate that must be requested by a receiver to guarantee a specific delay bound. In general, from an economic point of view, higher incoming C and D values lower the service quality, because a larger R is needed to achieve the same delay. The economic impact should be considered, for example, if an advanced QoS-oriented routing algorithm takes into account charges, it might be very important to have a quantitatively precise expression for this service degradation in order to value and compare different paths. The increase in service rate introduced by additional error terms (C_a,D_a) can be expressed as follows:

$$\text{let } X = \frac{b-M}{p-r} \tag{16}$$

$$R_a(C_a,D_a) = \frac{pX+M+C+C_a}{X+Q-(D+D_a)} - \frac{pX+M+C}{X+Q-D}$$
$$= \frac{C_a(X+Q-D)+D_a(pX+M+C)}{(X+Q-D-D_a)(X+Q-D)} \tag{17}$$

Because $R \geq r$, R_a is part of the q_C resource, the cost increase at each router is given by:

$$p(C_a,D_a) = a_C \cdot R_a(C_a,D_a) \tag{18}$$

Not considering the economic impact of error terms could lead to a situation where a service provider exports high C and D values in order to cause a higher reservation for R. Internally, however, the real C and D values can be used and a smaller reservation for R is needed to guarantee the end-to-end delay.

7.3 Slack Term

As denoted in the relevant research and standardization documents about Guaranteed service, e.g. [WC97,SPG97,GGP$^+$95], the slack term parameter S in a service request is intended to flexibly relax the resource requirements at intermediate routers.

Among other scenarios, this parameter can be used by first calculating the necessary service rate R depending on the desired end-to-end delay Q. Then, a higher service rate R^h is requested and the resulting difference of end-to-end delay is set as slack term parameter. The slack term can be “consumed” by an intermediate router for reducing both delay and rate requirements, depending on the type of scheduler. In case rate requirements are reduced, a bottleneck router installs a smaller service rate R^l according to the formula given in [SPG97]* and adjusts the reservation message, such that upstream routers only install R^l as well. The effects of using the slack term in such a way can best be explained by rewriting the delay formula for Guaranteed service [SPG97] as:

$$Q = \frac{(b-M)\cdot(p-\min(R_i))}{\min(R_i)\cdot(p-r)} + \frac{M}{\min(R_i)} + \sum_{i=1}^{n}\frac{C_i}{R_i} + D_{tot} \qquad (19)$$

for n routers, with R_i and C_i denoting the local service rate and error term at each router.

All downstream routers after a bottleneck router install the requested service rate R^h, which generates an economically unfortunate situation for the receiver. Considering (19), the smallest service rate installed along the path (in this case R^l) determines the pure end-to-end queuing delay. Having installed R^h at a number of routers only locally affects the additive delay component resulting from the local rate dependent error term C. Thereby, usability of the slack term in such a scenario largely depends on:

- the relation of delay introduced by C to pure queuing delay (mainly depending on p and b), and
- the relation of total upstream to total downstream amount of C

as given at the bottleneck router.

Effectively, the receiver pays for a higher service rate at some routers, but does not perceive the total utility from it. The costs can be expressed as follows:

$$\mathrm{cost}_{slack}(R^h,R) = a_C \cdot (R^h - R)$$
$$\text{at each router installing } R^h \qquad (20)$$

On the other hand, prices are lower at routers installing R^l:

$$\mathrm{save}_{slack}(R^l,R) = a_C \cdot (R - R^l)$$
$$\text{at each router installing } R^l \qquad (21)$$

When different service rates are installed at routers due to the basic slack term mechanism, it is very likely that costs exceed savings, because the resulting increase in pure queuing delay can only be recovered under pathological error term settings. Again, this economic impact should be considered when installing reservations using the slack term parameter.

Given (19), some suggestions can be made to extend the slack term mechanism. Often it could be advantageous to only locally install a lower rate and forward R^h instead of R^l to upstream routers. This would increase the maximum reduction of service rate and generate an even larger range of adaptability. In this case, the global minimum R (R_{min}) has to be transmitted in addition to the currently defined parameters and the slack term formula of [SPG97] has to be changed to:

$$S_{out} + \frac{(b-M)(p-R_{new})}{R_{new}(p-r)} + \frac{M+C_{toti}-C_i}{R_{out}} + \frac{M+C_i}{R_i} \le S_{in} + \frac{(b-M)(p-R_{min})}{R_{min}(p-r)} + \frac{M+C_{toti}}{R_{in}}$$

with $R_{new} = \min(R_{min}, R_i)$ and C_i and R_i denoting the local service rate and error term. (22)

The remaining variables have the same meaning as in [SPG97]. When the reservation is forwarded, R_{min} is set to R_{new}.

This idea can even be extended as follows. A router using the slack term sends a confirmation message containing its local C and D terms as well as its local service rate back to the receiver. Given this information, a receiver can optimize R^h when refreshing its reservation. In case of global price information, further optimization would be possible by setting a certain R^h at “cheap” links while setting R^l at “expensive” ones, e.g. a transatlantic link. In theory, this creates a linear optimization problem, but even if this can be solved, the necessary information exchange seems hardly be feasible with the current design of RSVP.

8 Summary and Future Work

In this paper, we examined price calculation for Integrated Services from a provider-oriented, profit-maximizing point of view. We established evidence that an internal price calculation formula should be linear for resource parameters and uniform across

---

*. Note that the slackterm formula on page 13 of [SPG97] should read:

$$S_{out} + \frac{(b-M)(p-R_{out})}{R_{out}(p-r)} + \frac{M+C_{toti}}{R_{out}} \le S_{in} + \frac{(b-M)(p-R_{in})}{R_{in}(p-r)} + \frac{M+C_{toti}}{R_{in}}$$

multiple service classes. Given these requirements, we suggested service refinements for IntServ classes and a mapping to uniform resource parameters. We demonstrated the applicability of this approach both for internal calculation and external price representation. Finally, we presented an economic analysis of certain characteristics of the Guaranteed service class with respect to such a calculation model.

Many issues remain open for further research. The problem of modelling interdependency between service classes respectively resource parameters requires a detailed understanding of usage patterns for future integrated services networks. The optimization problem presented in Section 6.1 is yet to be solved. Instead of a full theoretical analysis, it might be possible to build a periodically adaptive pricing system that measures actual demand and automatically derives demand estimations. In the technical area of Integrated Services, a refinement of service definitions is needed to precisely determine resource usage. When future services are introduced, this requirement should be kept in mind.

We are currently building a charging system incorporating mechanisms to exchange basic charging information when setting up RSVP flows [KSWS98] as well as the concepts presented in this paper.

References

[BCS94] Robert Braden, David Clark, and Scott Shenker. RFC 1633 - Integrated Services in the Internet Architecture: an Overview. Informational RFC, June 1994.

[CSZ98] Georg Carle, Michael Smirnov, and Tanja Zseby. Charging and Accounting Architectures for IP Multicast Integrated Services over ATM. In *Proceedings of 4th International Symposium on Interworking (Interworking’98), Ottawa, Canada*, July 1998.

[DVR98] Konstantinos Dovrolis, Maruthy Prasad Vedam, and Parameswaram Ramanathan. The Selection of the Token Bucket Parameters in the IETF Guaranteed Service Class. Technical report, University of Wisconsin-Madison, Madison, WI 53706-1691, USA, July 1998. Available from http://www.cae.wisc.edu/%7edovrolis/tokbuck.ps.

[FD98] Domenico Ferrari and Luca Delgrossi. Charging for QoS. In *Proceedings of 6th IEEE/IFIP International Workshop on Quality of Service, Napa, CA, USA*, pages vii–xiii. IEEE/IFIP, May 18–20 1998. Invited paper.

[FJ95] Sally Floyd and Van Jacobson. Link-sharing and Resource Management Models for Packet Networks. *IEEE/ACM Transactions on Networking*, 3(4), August 1995.

[FSVP98] George Fankhauser, Burkhard Stiller, Christoph Vögtli, and Bernhard Plattner. Reservation-based Charging in an Integrated Services Network. In *Proceedings of 4th INFORMS Telecommunications Conference, Boca Raton, Florida, USA*, March 1998.

[GGPR96] Leonid Georgiadis, Roch Guerin, Vinod Peris, and Raju Rajan. Efficient Support of Delay and Rate Guarantees in an Internet. In *Proceedings of ACM SIGCOMM’96*, pages 106–116, August 1996.

[GGP+95] Leonid Georgiadis, Roch Guerin, Vinod Peris, Raju Rajan, Dilip Kandlur, and Scott Shenker. A Case for Extension of the Guaranteed Delay Service Specifications, November 22nd 1995. Contribution to int-serv mailing list. Available from ftp://ftp.isi.edu/int-serv/int-serv.mail.

[GLV95] P. Goyal, S.S. Lam, and H. Vin. Determining End-to-End Delay Bounds in Heterogeneous Networks. In *Proceedings of 5th International Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV’95)*, April 1995.

[GSW95] Alok Gupta, Dale O. Stahl, and Andrew B. Whinston. A Stochastic Equilibrium Model of Internet Pricing. In *7th World Congress of the Econometrica Society, Tokyo, Japan*, August 1995.

[HSE97] Shai Herzog, Scott Shenker, and Deborah Estrin. Sharing the "Cost" of Multicast Trees: An Axiomatic Analysis. *IEEE/ACM Transactions on Networking*, 5(6):847–860, December 1997.

[KSWS98] Martin Karsten, Jens Schmitt, Lars Wolf, and Ralf Steinmetz. An Embedded Charging Approach for RSVP. In *Proceedings of 6th IEEE/IFIP International Workshop on Quality of Service, Napa, CA, USA*, pages 91–100. IEEE/IFIP, May 18–20 1998.

[KSWS99] Martin Karsten, Jens Schmitt, Lars Wolf, and Ralf Steinmetz. Cost and Price Calculation for Internet Integrated Services. In *Proceedings of Kommunikation in Verteilten Systemen (KiVS’99), Darmstadt, Germany*, March 2–5 1999.

[KVA98] Yannis A. Korillis, Theodora A. Varvarigou, and Sudhir R. Ahuja. Incentive-Compatible Pricing Strategies in Noncooperative Networks. In *Proceedings of the 17th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM’98)*, March 1998.

[Lei98] Brett A. Leida. Cost Model of Internet Service Providers: Implications for Internet Telephony and Yield Management. Master’s thesis, Massachusettes Institute of Technology, Cambridge, MA, USA, February 1998.

[MM97] Jeffrey K. MacKie-Mason. A Smart Market for Resource Reservation in a Multiple Quality of Service Information Network. Technical report, University of Michigan, September 1997. Available from http://www-personal.umich.edu/%7ejmm/papers/reserve3.pdf.

[MM98] Jeffrey K. MacKie-Mason. Multi-Agent Systems, Mechanism Design, and Institutions. In *First International Conference on Information and Computation Economies (ICE-98)*, 1998. Invited Talk.

[MMV95] Jeffrey K. MacKie-Mason and Hal R. Varian. Pricing the Internet. In Brian Kahin and James Keller, editors, *Public Access to the Internet*, pages 269–314. MIT Press, Cambridge, 1995.

[MMV97] Jeffrey K. MacKie-Mason and Hal R. Varian. Economic FAQs About the Internet. In Joseph Bailey and Lee McKnight, editors, *Internet Economics*. MIT Press, Cambridge, 1997.

[MV91] Bridger M. Mitchell and Ingo Vogelsang. *Telecommunications Pricing, Theory and Practice*. Cambridge University Press, 1991.

[PSC98] Kihong Park, Meera Sitharam, and Shaogang Chen. Quality of Service Provision in Noncooperative Networks: Heterogeneous Preferences, Multi-Dimensional QoS Vectors, and Burstiness. In *Proceedings of First International Conference on Information and Computation Economies (ICE-98)*, pages 111–127, 1998.

[RPH98] Michael H. Rothkopf, Aleksandar Pekec, and Ronald M. Harstad. Computationally Manageable Combinational Auctions. *Management Science*, 44(8), August 1998.

[San98] Tuomas Sandholm. Winner Determination in Combinatorial Auctions. In *First International Conference on Information and Computation Economies (ICE-98)*, 1998. Invited Talk.

[SCEH96] Scott Shenker, David Clark, Deborah Estrin, and Shai Herzog. Pricing in Computer Networks: Reshaping the Research Agenda. *ACM Computer Communication Review*, 26(2):19–43, April 1996.

[SFY95] Jakka Sairamesh, Donald F. Ferguson, and Yechiam Yemini. An Approach to Pricing, Optimal Allocation and Quality of Service Provisioning in High-Speed Packet Networks. In *Proceedings of the 14th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM’95)*, pages 1111–1119. IEEE Computer Society Press, June 1995.

[SPG97] Scott Shenker, Craig Partridge, and Roch Guerin. RFC 2212 - Specification of Guaranteed Service. Proposed Standard, September 1997.

[WC97] Paul White and Jon Crowcroft. Integrated Services in the Internet: State of the Art. *Proceedings of IEEE*, 85(12):1934–1946, December 1997.

[WPS97] Qiong Wang, Jon M. Peha, and Marvin A. Sirbu. Optimal Pricing for Integrated-Services Networks with Guaranteed Quality of Service. In Joseph Bailey and Lee McKnight, editors, *Internet Economics*. MIT Press, 1997. Available at http://www.ece.cmu.edu/%7epeha.

[Wro97a] John Wroclawski. RFC 2210 - The Use of RSVP with IETF Integrated Services. Informational RFC, September 1997.

[Wro97b] John Wroclawski. RFC 2211 - Specification of the Controlled-Load Network Element Service. Proposed Standard, September 1997.

[Zha95] Hui Zhang. Service Disciplines for Guaranteed Performance Service in Packet-Switching Networks. *Proceedings of the IEEE*, 83(10), October 1995.