Charging for Packet-switched Network Communication – Motivation and Overview –

Martin Karsten[1], Jens Schmitt[1], Burkhard Stiller[2], Lars Wolf[1,3]

1: KOM – Industrial Process and System Communications
Darmstadt University of Technology
Merckstr. 25, 64283 Darmstadt, Germany

2: TIK – Computer Engineering and Networks Laboratory
Swiss Federal Institute of Technology, ETH Zürich
Gloriastrasse 35, CH-8092 Zürich, Switzerland

3: GMD IPSI
German National Research Center for Information Technology
Dolivostr. 15, 64293 Darmstadt, Germany

{Martin.Karsten,Jens.Schmitt,Lars.Wolf}@KOM.tu-darmstadt.de
stiller@tik.ee.ethz.ch

Abstract Packet-switched network technology is expected to form the integration layer for future multi-service networks. The trend to integrate voice and data traffic can already be observed in the Internet and is expected to carry on until the full integration of all media types is achieved. At the same time it becomes obvious that the business model employed for current Internet usage is probably not sustainable for the creation of an infrastructure suitable to support a diverse and ever-increasing range of application services. Currently, the Internet provides only a single class of best-effort service and prices are mainly built on flat-fee, access-based schemes. However, the large and increasing differentiation of demand for quality of transmission services creates the necessity to differentiate service provision and, consequently, charging for network communications. In this article, we survey relevant service models and motivate the need for appropriate charging mechanisms. We give an overview on requirements for a charging system, its technical components, and issues for future business models that might by employed by service providers.

Keywords Packet-Switched Networks, Quality of Service, Pricing, Charging

1 Introduction

The current situation for communication networks is characterized by the fact that different network technologies are used to carry data of different applications on dedicated network infrastructure. Examples are given by the telephone network, cable-TV networks, radio broadcast, and leased-line service for mission-critical data transfers. The Internet as it currently exists, is designed for elastic data applications [She95b]. Little real-world experience is available on reliably integrating multiple different application services in a single network in a ubiquitous and commercial fashion, even though, theory tells us that service integration is an efficient manner for providing communication services [She95a].

Employing a single basic network infrastructure to create such an integrated services network promises a potential for cost reduction, which in total is much larger than tuning different technologies for different types of applications. Many researchers assume packet-based communication to render the most flexible and, thus, most appropriate alternative to integrate different link technologies and provide a unique bearer service. Hence, the major challenge is to support a broad and diverse range of applications. There are many proposals (cf. Section 2) to enable these kinds of usage, for example by providing multiple service classes at the network layer.

The current Internet is characterized by the existence of a single class of best-effort service, (assumed) cooperation among users, and fairness built into end-systems due to TCP's congestion control algorithm [Ste97]. For a future commercial environment, some of these assumptions might have to be revised, in particular, it might not be suitable to rely on cooperation amongst users and fairness enforced by appropriate protocol implementation in routers or end-systems. Instead, no matter what kind of service differentiation is carried out at the network level, effective feedback signals have to be generated to protect the network from arbitrary usage. On one hand, such a feedback signal is supposed to be conveyed by pricing, setting a certain price for a unit service used. Appropriate pricing of network communication provides incentives for reasonable usage of resources. Correspondingly, the notion of fairness of network operation has to be changed from max-min fairness ("everyone is equal everywhere") [BG87] to a notion of proportional fairness [Gar95] relative to the amount of charges one is willing to pay. On the other hand, a market and competition mechanism is useful to provide users with the best and most inexpensive level of service, while creating incentives for network providers to supply more resources when there is sufficient demand.

This article is intended to survey current state of the art and future issues for building a commercial integrated network based on packet-switching technology. In the next section we begin by briefly reviewing the relevant aspects of service provisioning and give an overview of different proposals in this area. In Section 3, we derive the motivation for differentiated charging and introduce and explain fundamental terminology. Requirements for a charging system are discussed in Section 4. Section 5 presents generic technical components of a charging architecture. In Section 6, we discuss the complex issue of finding appropriate business models for communication providers in the presence of competition and rapid technological advances. This paper concludes in Section 7 with a summary and an outlook to further research issues.

2 Service Provision

Network services can be provided by a spectrum of different technologies which can be mainly distinguished by their trade-offs between complexity and capabilities. Those technologies are built upon different models of Quality of Service (QoS). A *QoS model* consists of the following macroscopic facets (related to [Bra97]):

- **Scope** defines the logical distance over which a service model is provided.
- **Granularity** defines the smallest unit which is treated individually by a service model.
- **Time Scale** defines the granularity in time on which services are being provided.
- **Control Model** defines the entities which exert control over the network and how they do it. As extreme cases control could either be located exclusively in the network or in the end-systems, with a continuum of hybrid forms in between. regulation

QoS models apply different tools in order to achieve their respective goals:

- **Network Design and Engineering** deals with the proper setup and maintenance of network equipment based upon experience, expert knowledge, heuristics or formal optimization methods [Ker93]. Sometimes this is also called *provisioning*.
- **Traffic Engineering** is concerned with distributing the traffic for a given network by mechanisms as, e.g., explicit routing capabilities [RVC99] or QoS-based routing schemes [ZSSC97].
- **Signalling and Admission Control** is an integrated set of mechanisms which builds upon a session paradigm where users of the network signal their requirements explicitly and the network consults local admission control modules to grant or reject those requests. Examples of proposed signalling protocols for the Internet are the Resource Reservation Protocol (RSVP) [BZB^{+}97] and the Stream Protocol (ST2+) [DB95], while proposed admission control procedures are either parameter- [PG93] or measurement-based [GK97,BJS99]. With-

out per-flow admission control, only statistical QoS assurances can be given on a per-packet basis.

- **Packet Scheduling** is concerned with the decision which packet to send next on a given link if there is a number of buffered packets waiting for service [SV98]. This decision of course has a major impact on the QoS experienced by a packet since the queuing delay constitutes a large portion of the total end-to-end transfer delay.
- **Traffic Policing/Shaping** deals with forming traffic to an either negotiated or at least advertised level at the edges of networks or between distinct network elements in order to ensure a controllable load of the network. Example mechanisms in this area are the well-known leaky or token bucket traffic envelopes [Tur86,Cru91].
- **Adaptiveness** is the capability of end-systems to react upon congestion in the network by evaluating signals from the network. These signals can be either implicit, e.g., the loss behaviour of the network, or explicit, e.g., by a so-called Explicit Congestion Notification (ECN) [RF99]. Dynamic and possibly congestion-based pricing of network services is also a form of network signal proposed for managing QoS [MMV95,GK99b].

These tools have different time-scales and, thus, not all these tools are suitable for all different kinds of QoS models. However, in general a combination of those tools needs to be applied to implement a certain QoS model. For example, proper network design and engineering is certainly a prerequisite to the successful operation of any QoS model. The QoS models mainly differ in how much emphasis is put on each element in their combination of tools, which reflects different assumptions on how powerful the different components are assessed.

Proposed QoS models in the Internet arena are listed in the following and classified according to the above mentioned aspects.

RSVP/IntServ. This model [BCS94] is composed of RSVP which represents a specific signalling protocol and service classes defined by the Integrated Services (IntServ) architecture [Wro97,SPG97]. Its scope is to provide end-to-end services, and its granularity is determined on a per-flow basis, i.e., it is very fine-grained. Through the RSVP protocol the concept of sessions is introduced which determines the unit of time-scale of that model. The control loop of the RSVP/IntServ model tends to be network-centric in the sense of offering fairly advanced services inside the network among which applications may choose from. A usual counter-argument against this QoS model is based on the resulting complexity for network elements.

The most important tools to implement the RSVP/IntServ model are signalling and admission control obviously, however it also depends upon a sensible network design and engineering in order to keep the blocking probability for sessions low. Furthermore, policing and shaping is required to keep reserved traffic in its negotiated borders.

DiffServ. While still evolving [BBC$^+$98], the Differentiated Services (DiffServ) model's scope is rather inter-domain, based upon the peering between domains. It outlines a framework which allows for bilateral contracts by Service Level Agreements (SLA) at such borders. Currently, the DiffServ proposals only deal with different flavours of forwarding behaviours inside network elements, the so-called PHBs (Per-Hop Behaviours) [HBWW99,JNP99]. It is assumed that by concatenating PHBs, it is possible to build sensible services, thereby allowing for an end-to-end scope in result. However this is not a necessary direction of evolution for DiffServ. The PHBs operate on traffic aggregates, or so-called BAs (Behaviour Aggregates), and thus its granularity is fairly coarse-grained. Similarly, the unit of time-scale is expected to be long-termed since SLAs should be rather static, although with the addition of dynamic SLAs by the introduction of signalling protocols, the time-scale could decrease [BC99]. As with the RSVP/IntServ model, the control model is fairly network-centric, but for some of the PHBs defined it is necessary for end-systems to adapt themselves to the network state.

The tools upon which DiffServ builds are mainly network design/engineering and traffic policing or shaping. However, an introduction of dynamic SLAs will make it move towards an emphasis of signalling and admission control mechanisms, also.

Over-provisioned Best-effort. This model argues for a continuation of the current operation of the Internet in a best-effort manner. The underlying assumption is that over-provisioning network resources is both possible and sufficient to sustain the single service nature of the current Internet. The scope of that model is end-to-end in nature since all the intelligence is located in end-systems. Since there is no state in the network and all traffic is treated at the same granularity, it is as coarse-grained as possible. The time-scale of this model is also very large and essentially equal to the length of one capacity planning cycle. Since end-systems are the only intelligent units in the network the control model is end-system-centric.

Certainly, the most important tool applied by that model is that of network design/engineering in order to always provision for a super-abundance of network resources. However, in periods of scarcity of resources this model relies on the adaptiveness of end-systems to such presumably transient situations.

Price-controlled Best-effort. This is not a single proposal but a notion of several authors [MMV95,KMT98,CP99] who feel that pure over-provisioning is not sufficient without an additional means of signalling besides packet loss. This additional signal is a per-packet price that may depend on the internal state of the network, e.g., its congestion level. However, some authors even propose a semi-static approach with fixed but differentiated prices per packet [Odl99]. With respect to its properties this model is very similar to the pure over-provisioned best-effort model, however, its time-scale is related to the frequency of price announcements and, due to the ability to set prices, the network is not as passive as for that model.

With regard to the tools that are applied by that model it also has to be noted that it heavily relies on the combination of network design/engineering and the adaptiveness of end-systems. Furthermore, it is crucial for correct operation that the end-systems' or users' sensitivity to pricing signals can be estimated. However, there are open questions about potential time-gaps between pricing signals and adaptation at the end-systems, as well as the prerequisite of global consensus between all participants in such a model.

Those different proposals address different application requirements and, therefore, different business models (cf. Section 6). None is necessarily right or wrong. This is especially based on the observation that no large-scale trial has been performed to achieve real-life data and results. Complicatingly, the combination of technical means as outlined above and an economic-based business model integrates a set of not yet fully understood factors (such as packet-switched, connection-less networking technology and the extremely high pace of network and customer growth) of a new and quite rapidly emerging Internet services market.

3 Terminology and Motivation for Charging Network Services

For each of the QoS models discussed in Section 2, there is a different motivation whether and how to carry out charging for network services. As well, each model can be combined with a pricing scheme which depends on the unit service, which has been determined to be pricing relevant. While in principle each of these combinations (QoS model and pricing scheme) seems possible, we argue that certain combinations are more reasonable than others. This is due to the fact that the diverse tools for QoS employ different notions of resource usage. Furthermore, many different optimization directions exist, e.g., service providers like to maximize their revenues, end-customers tend to minimize their communication costs, and content providers intent to offer quality-content while thrusting aside networking costs, and not every combination allows for an adequate optimization. In this section, we begin by introducing basic terminology for the business of commercially providing network services. Then, we enumerate and assess the combinations of QoS models and pricing schemes. As well, we briefly explain the different rationales behind charging for combinations that we assess as reasonable.

3.1 Terminology

In order to allow for an unambiguous discussion, we introduce basic terms and their meaning in the context of packet-switched networking. In general, the business field of network services is characterized by the following aspects:

- high fixed costs (installation and maintenance of infrastructure),
- low variable costs,
- fixed capacity, and
- non-storable resources and products.

In business economics, the appropriate management theory for such a business field is given by *Yield Management* (cf. [Lei98] and references herein). Under these characteristics, it is more appropriate to differentiate prices according to variations in demand, instead of doing full cost calculation for prices. Essentially, the *marginal return*, which is given by the difference between sales price and variable costs is considered to be the primary variable to take into account. The goal is to optimize the sum of marginal returns over a certain investment cycle. This sum must exceed the respective investments in order to allow for profitable business.

The *resources* of a packet-switched communication infrastructure are given by computing power and buffer space of switching systems and capacity of transmission lines. Capacity of a transmission line is useless, if it cannot be utilized fully by the feeding switching system. Computing power mainly determines the amount of packets that can be serviced and the amount of flows that can be handled in case of flow-based QoS models. Additionally, it determines the possible levels of service differentiations. Therefore, the resources that are to be assessed for price calculation are given by:

- flow setup overhead (if existing),
- packet rate (and its schedulability),
- traffic bandwidth (not necessarily constant), and
- buffer space.

In the above list, the terms *packet rate* and *traffic bandwidth* are listed separately to distinguish between different units of measurement for transmission capacity. The number of packets that can be handled is mainly determined by computing power of a switching system whereas traffic bandwidth is limited by the overall link bandwidth and throughput capacity of a switching system. An additional important input parameter for capacity dimensioning and, thus, for price calculation, is given by the access bandwidth of users or adjacent providers. This parameter eventually determines the maximum resource quantity of service invocations.

In the context of yield management, since real variable costs basically do not exist or are extremely limited in nature, the term *costs* per service invocation can be used to characterize so-called *opportunity costs*, i.e., lost revenue because resources are bound and cannot be sold otherwise. Consequently, resource usage and, therefore, consumption can be considered as the main cost factor.

On one hand, the *price* for a service, for a service invocation, a certain quality level of the service, or in general for a unit service depends on costs, demand, and general marketing considerations. On the other hand, prices provide feedback signals to users and, thus, for their part influence demand and usage. An important distinction must be made on whether prices are set ahead of time (*fixed price*) or determined and potentially changed during service invocation (*variable price*). For the latter a good example represents an *auction* [Vic61], [Var96] with continuous price variability considered as repeated incarnations of the auctioning process.

If prices are set, taking into account the network's utilization level, they are termed as congestion-based prices or load-sensitive prices. These types of prices are considered of being responsive to the overall usage of resources, as determined above. Their main advantage lies in the capability to provide feedback signals from the network to the user while offering incentives at the same time to use a service at a certain point in time or to back off during congested and higher utilized network periods. A *pricing scheme* describes a particular choice from all the possibilities presented above and is applied to unit services offered from a communication service provider.

As mentioned, prices are associated with units of service. Therefore, these units need to be accounted for, traditionally performed on a per-call basis over time. However, in packet-switched networks, the accounted for information may encounter a huge number of different parameters, e.g., number of packets sent, duration of a communication, number of transactions performed, distance of the communication peer, number of hops traversed, or bandwidth used. Depending on the protocol layer applied for this *accounting* task, only a subset of accounted for parameters are useful. In general the *accounting record* determines the container for collecting this information. These records and their special appearances depend on the networking technology used, such as N-ISDN, ATM, Frame Relay, or IP. They can also be created for application services, for example, the *call data record* is being used for this purposes in H.323 IP telephony. Furthermore, the Real-time Flow Measurement working group within the IETF investigates appropriate accounting mechanisms [BMR97].

Once these accounting records are collected and prices are determined in full pricing schemes on unit service, e.g., encompassing different quality levels for services or service bundles, the data for an invoice need to be calculated. The process of this calculation is termed *charge calculation*, performing the application of prices of unit services onto accounted for records determining the resource consumption. Thus, the charging function transforms mathematically unequivocal technical parameter values into monetary units. These units need to be collected, if they appear at different locations in the given networking environment, and are stored in *charging records*. Of course, accounting as well as charging records determine a critical set of data which need to be secured to ensure its integrity when applied to calculate monetary values or when used to compute an invoice's total.

The process of consolidating charging records on a per customer basis and delivering a certain aggregate of these records to a customer is termed *billing*. The basic task of collecting these charging records requires sufficient protocol support, including authentications, to allow for counterfeit-proof invoice computation. This aggregation of monetary values may be performed on a daily, weekly, monthly, or some other accepted period of time. The invoice or bill summarizes a number of contracted details on the parameters originally collected in the accounting records. Additionally, it determines the amount of money to be paid from a customer to the service provider which may be delivered traditionally on paper or in an electronic fashion. Furthermore, it may determine the method of *payment* as well. Since the payment defines the method of how the exchange of money between buyers and sellers will be performed, traditional methods may be used or advanced electronic payments schemes may be applied.

Last not least, there remains a single technical prerequisite for identifying and collecting accounting data. This process is called *metering*. Based on existing technical equipment in operation, the metering tasks identify the technical value of a given resource and determine their current usage. If possible, metering can be tied to signalling events [CSZ98, FSVP98, KSWS98]. Otherwise, it may be performed regularly, e.g., every ten seconds or every hour, it may be stimulated on other external events, such as polling requests, or it may be performed according to some statistical sampling scheme. In that case, it is closely related to network monitoring. The IETF's Management Information Bases (MIB) for switched networks [WLRW99] and the Simple Network Management Protocol (SNMP) architectural framework [HPW99] may provide a means of keeping monitored data.

Finally, concerning this overall terminology discussion, sometimes the terms pricing, charging, or billing are used to represent the complete process of detecting the specific usage of a service, its pricing, accounting, charge calculation, and billing as defined above [SFPW98]. To employ a precise and unambiguous notion throughout the paper, these separate terms are used in their appropriate meaning, while *charging* is used to determine the complete process.

3.2 Pricing Schemes for QoS Models

Table 1 gives an overview of which pricing scheme fits with which QoS model. The notation is as follows. The plus sign (+) denotes a good fit between QoS model and pricing scheme. A zero (0) denotes that it seems questionable to us whether such a combination is reasonable, and a minus sign (-) denotes that this combination of QoS model and pricing scheme is likely not to be viable.

	Flat fee	Resource-based (Fixed prices)	Resource-based (Variable prices)
Over-provisioned Best-effort	+	-	0
Price-controlled Best-effort	-	-	+
DiffServ	-	+	+
IntServ	-	+	+

Table 1: Combinations of QoS Models and Pricing Schemes

In Table 1, *flat fee* denotes the current access-based pricing scheme of the Internet. The term *resource-based* denotes a pricing scheme that is individually based on the amount of resources used for a service invocation or service usage. This could be metered, for example by measuring traffic or pricing for a resource reservation task. These two terms denote extreme ends of a spectrum of possible pricing schemes. Arbitrary combinations and hybrid approaches seem possible.

The nature of over-provisioned best-effort service is such that no service discrimination is possible, and hence, price discrimination is not appropriate[1]. Although best-effort services have been used in combination with fixed per-packet prices, we do not consider this a useful alternative, because fixed prices do not represent the resource consumption of best-effort communication. In case best-effort services are combined with resource-based pricing and variable prices, this basically resembles price-controlled best-effort service. In general, it seems doubtful, whether this QoS model is capable of providing that kind of service that is needed for differentiated application demands. Even when assuming an ever increasing amount of transmission resources at constantly decreasing prices, a situation of super-abundance can only exist in relation to a certain amount of aggregated demand. Such a system must be kept flexible with regards to requests from end-users to attract widespread usage. Nevertheless, for reliable operation, it must be ensured that aggregated demand does not exceed an acceptable level, i.e., that network load actually remains to be low. To combine both requirements, some kind of dynamic access control is needed (1) to ensure proper and controllable consumption of resources and (2) to account for any premium service usage.

---

[1] The price may vary according to the access bandwidth provided to the customer. Still, this determines a flat fee for the customer, however, the service provider offers a limited set of resource-based pricing schemes.

For price-controlled best-effort service, appropriate pricing and responsiveness of end-systems to price signals is the crucial management aspect. Because of the responsiveness of end-systems, per-packet charges provide a mechanism for dynamic access control. Under the assumption of stable price-demand patterns, it is possible to dimension capacity such that reliable operation and QoS assurances can be met at least statistically. However, since performance predictability can only be given under certain restrictions (cf. [PSC98]), such a service cannot provide the exclusive technology for the overall network infrastructure. Furthermore, prices are inherently variable in order to fulfil their functionality as congestion signals. It has been suggested to combine such a basic service with higher level entities, which act as trading or insurance brokers to remove price fluctuations or improve QoS predictability. However, it may add a significant complexity to the overall system to implement such brokers and fine-grained interactions between them, if the frequency of these interactions reaches a certain limit. Future investigations need to carefully design, simulate, and implement such systems to provide evidence for their feasibility. To this end, we consider price-controlled best-effort as an alternative implementation choice for certain service classes that do not specify hard QoS guarantees, e.g., similar to the Controlled Load service class [Wro97].

In the DiffServ and IntServ model, resources are reserved or engineered according to the requested service invocations. Independent of actual service implementation, some kind of admission control has to be executed on service requests in order to guarantee reliable and predictable transmission quality, as specified in the respective service classes. Since resources are allocated (more or less exclusively) to service requests and, therefore, these resources are not available to others, charging has to be done resource-based in order to keep the demand at a commendable level and to avoid the *tragedy of the commons* phenomenon [Har68]. While these technologies incur a relatively high complexity at the technical level of service provision in the network (IntServ higher than DiffServ), they also provide the most sophisticated interfaces to network management and users (again, IntServ more than DiffServ). Consequently, the additional complexity for providing a large range of different application services and pricing and charging for these services is lower than for price-controlled best-effort approaches. Initial proposals for appropriate pricing models can be found in [MM97,WPS97,SK97,KSWS99].

3.3 Telephony vs. Internet Charging

Charging communication services in general, such as network, transport, or application services, does not define a new area as such, however, charging of Internet services has not been performed explicitly visible for end-users or not at all. Therefore, a concept for introducing charging into the Internet is required. As described above, charging becomes a necessity, if different services are offered in a multi-service network. Simply applying existing solutions for, e.g., telephony charging to the Internet does not work, however.

One of the important reasons is due to the connectionless vs. connection-oriented type of network. Established telephony charging systems do access connection state information about caller and callee, their locations and connection time. However, an Internet charging system must be able to account for connectionless traffic as well, where state or other information may not be available. Besides this end-user’s point of view, charging between providers cannot be based on connection information only, as well. Furthermore, traffic specifications and models vary between the traditional telephone and Internet traffic. This is visible as well in a wider range of communication parameters for the Internet, such as bandwidth, delay, inter-packet times, or error rates. Peering agreements including these parameters set forth a new degree of complexity, which is still investigated today. Finally, the Internet differs from traditional telephony due to its inherent multi-service functionality. Especially this service discrimination functionality requires Internet charging systems to keep track of more information, compared to the telephony charging system.

4 Requirements for Commercial Communication Networks

In this section, we enumerate a set of requirements that are very likely to apply to commercially funded and charged network communication. These requirements can be classified along the following aspects.

- For charging systems, a number of criteria arise from current practice and user expectations in the area of product liability and consumer protection in general.
- The business task of operating a communication network must be sustainable and profitable enough to attract the necessary investments in infrastructure.
- General objectives for a charging system, for example in terms of flexibility and efficiency, have to be met.
- Objectives for overall network operation are raised by introduction of new services and charging thereof.

Similar requirements have been established and discussed, e.g., in [FD98, KSWS98, MP99].

4.1 User Requirements

Predictability of Charges. Users want to be able to predict the costs of using a particular application, which include the expenditures for the communication services induced by this application. Therefore, an exact a priori specification of communication charges would be desirable. However, if this requirement cannot be fulfilled, a set of weaker demands can be sufficient. First, a user should be able to roughly estimate his charges. Such an estimation does not need to be exact but should give at least a rough feeling to the user – similar like the knowledge that an international phone call of some minutes duration costs more than a dollar and not just a few cents. Second, a worst-case price should be known. Finally, it must be prohibited that a user is charged a higher price than previously announced, without giving his explicit approval.

Transparency and Accuracy of Charging. To find out how much is spent for which application and what are the reasons for this, users need the ability to determine the costs of a particular session, e.g., if an application uses several flows, the costs for each of these should be stated explicitly. Furthermore, for some users it might also be of interest to see where inside of the network the major charges are caused. This may give them information to switch to a different provider in future. Detailed per-session information about charges can also be used to decide whether a certain service and its quality offer good value for the price. Since not all users are interested in such details, each user must be able to decide how much information should be given.

Convenience. Charging components should not make the usage of communication services much more difficult. The charging mechanisms themselves as well as the final bill based on the information gathered by the charging system must be convenient for its users. Hence, it must be possible for users to define “standard charging behaviour” for their applications so that they are not bothered with details during the start up of an often used application. On the other hand, they should be able to change such a description easily to have control over their expenditures, e.g., changing spending caps. Furthermore, most users want to have as few separate bills as possible, i.e., have contracts and according business procedures with only one provider.

4.2 Provider Requirements

Technical Feasibility. The charging approach and its mechanisms must be implementable and operable with low effort. Otherwise, if it becomes too complex, the costs for the charging mechanisms might be higher than their gains. A set of real-life user trials needs to be performed to assure any of these statements. The added overhead for communication due to additional information transmitted between senders, network nodes, and receivers, and also for processing and storage purposes especially in network nodes, e.g., to keep and manipulate charging information, must be as low as possible [FSVP98]. In addition, the introduction of scalable and low effort security mechanisms is essential for any type of counterfeit-proof charging records and billing data.

Variety of Business Models. The business of providing network service over packet-switched networks must be sustainable and profitable to attract the necessary investments into the infrastructure. It is unlikely to expect all service providers to adopt exactly the same business model and strategies. Therefore, charging mechanisms must be flexible enough to support a large variety of business models and interoperate between multiple network domains employing different models. As well, a charging system must be flexible enough to handle different pricing strategies, for example during peak and off-peak times. The issue of choosing an appropriate business model is discussed in Section 6.

4.3 System Requirements

Flexibility. When information is transmitted from a sender to one or several receivers, the flow of value associated with this information can be (1) in the same direction as that of the data flow, (2) in the opposite direction, or (3) a mixture of both because both sides benefit from the information exchange. For example, in the first case, the sender transmits a product advertisement, in the second case, the receiver retrieves a movie for playback, and in the third case both sides hold a project meeting via a video-conference system. To support these different scenarios, a charging architecture must provide flexible mechanisms to allow the participants in a communication session to specify their willingness to pay for the charges in a variety of manners. Senders must be able to state that they accept to pay for some percentage of the overall communication costs or up to a specified total amount. Similarly, receivers may state what amount of costs they will cover. Additionally, charging mechanisms must allow to flexibly distribute communication charges among members of a multicast group. A number of cost allocation strategies can be found in [HSE97].

Fraud Protection and Legal Security. One of the most important issues demanded by participants is protection against fraud, i.e., that they do not have to pay for costs they have not incurred and that no one can misuse the system. The fear of users is that a provider may cheat or that other users may use their identity or derogate from them in any other way. Providers want to be sure that users indeed pay for the used service. A prerequisite against fraud is technical security, such that users cannot damage, misuse or intrude the provider’s communication systems. Finally, legal security denotes the demand that in case of a failure, there is enough information to determine responsibility for it.

4.4 Network Operation Requirements

Stability of Service. When a particular service with a certain quality has been agreed upon by the user and the provider, it must be ensured that the service indeed is delivered to the user. Hence, an exact definition of “quality assurance is met” is needed. On the other hand, users must be able to estimate the impact of such quality goals on their applications, hence the definition must not be too complex. For example, multiple users start a video conference application, thus they likely request a communication service with a specified bandwidth and delay. If the provider assures to deliver this service, the users expect no quality degradation and a very low probability of service disruption during the conference. In case of quality degradation or service disruption, an appropriate refund mechanism must be applied, which largely depends on the type of application, and hence, should be negotiated during set up of the communication service.

Reliability of Service. In order to provide the infrastructure for an integrated packet-switched network, service availability must be very reliable. Current telephone networks are designed to keep the blocking probability in the order of 10^{-4}. Similar requirements are likely to apply to integrated services networks as well. To assure such a low blocking probability, even during peak hours, significant effort in the area of network and traffic engineering is necessary, which in turn must be accompanied by appropriate business calculation. A slightly different situation exists in case of per-packet QoS guarantees without explicit flow admission control. In that case, the notion of blocking probability might be replaced by reliability of service measured in terms of probability that the promised level of QoS is violated.

5 Technical Components of a Charging System

Several technical components are needed to provide for an overall charging system. As illustrated in Figure 1, the general scenario contains various interconnected communication service providers. Each provider has a network consisting of routers and network links between them, accounting systems, and a billing system. Metering systems are components inside of the networks. They can be independent components or be combined with routers. In either case, they generate accounting information (base accounting records) which are gathered and accumulated in accounting systems. The accounting systems in turn forward the accumulated and perhaps abstracted accounting information through a charge calculation function towards the billing system. The charge calculation translates the accounting information into charging records, hence, it maps the resource-oriented information from the accounting systems into monetary values. The billing system uses these values to prepare the bills to be sent to users/subscribers. Within the charge calculation, and perhaps the billing system as well, any discounting strategies, marketing-driven pricing schemes, or simply fixed prices are applied.

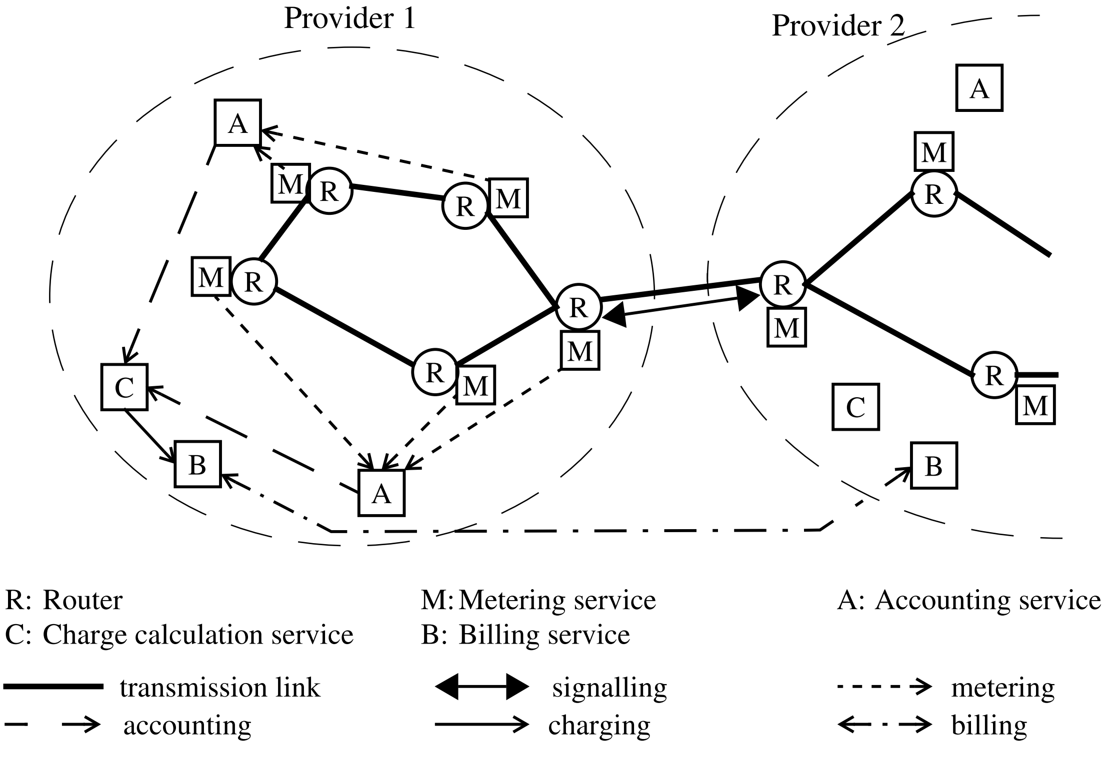

Figure 1: Charging Components in Provider Networks

Information exchange between providers occurs on the level of the billing systems, where inter-provider invoices are exchanged, and routers. Instead of performing absolute billing between interconnected providers, they can also offset their claims against each other. A huge set of peering agreements and settlement schemes exist for todays Internet Service Providers, however, they are defined in a quite static manner and do not allow for immediate responses to bandwidth bottlenecks or further end-customer demands. Further inter-provider information exchange happens as part of specific protocol processing as defined in the QoS model applied, e.g., for resource reservation purposes such as using RSVP or inter-Bandwidth-Broker communication, where messages are exchanged between the boundary routers of neighbouring providers.

In existing billing systems, the setting of prices, the function of charge calculation, and the billing itself is integrated, even additionally combining the maintenance of service classes, user profiles, customer data, identities, and banking account data. Although the above mentioned steps still can be distinguished clearly, they are almost completely centralized within a single system, except for the accounting task. Future billing systems need to be able to integrate a variety of different charging records, even from different communication providers or content providers, since one-stop billing determines an end-customer's demand [SCEH96].

The metering systems imply the role to deliver the basic information used for charging. They have to detect the resource 'usage' by subscribers – which can be the actual usage for transport and which also can be the reserved but finally unused capacity. For reserved resources, information available within routers or Bandwidth Brokers should be used. Therefore, the Protocol Data Units (PDU) exchanged among these entities may contain (policy) objects which describe announced prices and according charges [Her99,KSWS98]. For traditional best-effort data traffic, no such explicit information about resource demand and usage is available. Therefore, and also for potential volume-based charging for reserved traffic, metering services are necessary which measure and count the transferred data volume. These units can be part of routers or additional devices attached to LANs or to metering ports of routers. Of course, depending on the performance requirements of such a router or link, using separate devices can offer further advantages besides performance in the area of reliability, division of responsibilities, etc.

Of course, the basis for a solid and acceptable pricing model is determined by various issues, such as the technical possibilities of the metering task, the performance required by this task, and its granularity of information collection. Only the data being metered may be used for the accounting record and the unit service being priced, which in turn determines the maximum billing granularity achievable. Certainly, the degree of granularity of these information may lead to a severe technical inefficiency for metering data to apply usage-sensitive pricing methods, if care has not been taken to reduce this data set to the basic essentials only.

These technical issues have been studied mainly in the context of RSVP and IntServ, so far. According architectures and components have been discussed by, e.g., [FSVP98, KSWS98, CSZ98]. These approaches extend the RSVP protocol to distribute charging and accounting information among routers, respectively. Additionally, protocol mechanisms to forward usage information from the metering systems to the accounting systems are required.

The conceptual separation between the various types of components and their interactions can be continued as shown in Figure 2. Usage information from the metering services is used to generate pricing information which is injected back into routers/metering points. It might be important that pricing data is available for the information exchange within the network layer, therefore, it is not sufficient to distribute this information only between policy nodes. In addition, the customer identification is required already during the metering task, latest at the accounting stage to be able to identify the person responsible for the later payment of the service.

By now, there are many technical aspects which have not been studied in sufficient detail in the literature. Furthermore, the interplay of the various approaches to provide quality of serv-

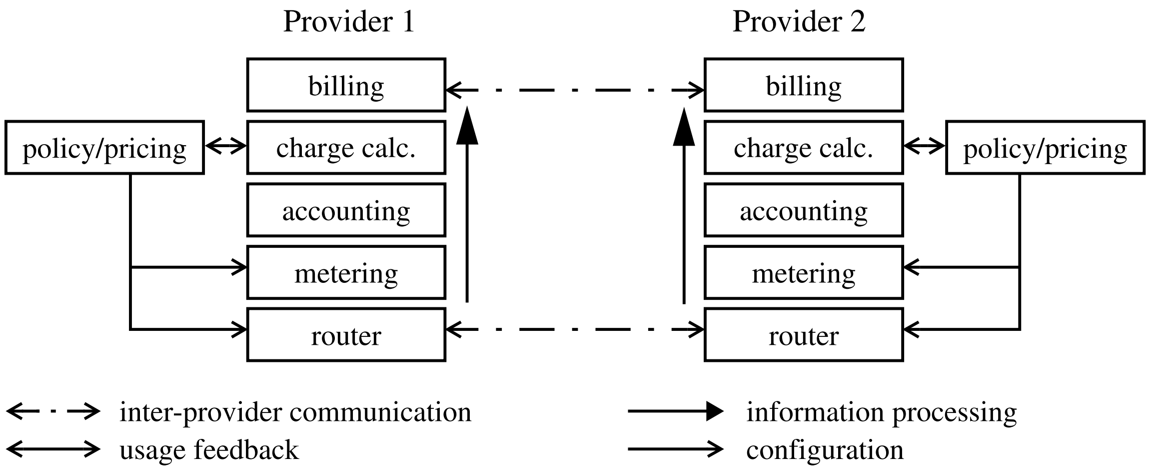

Figure 2: Conceptual View of a Charging System

ice makes the situation much more complicated. The discussion of all these issues is out of the scope of this article. As an example, we briefly describe the problems which already occur for relatively well established communication mechanisms such as IP multicast. The treatment of IP multicast is problematic in general due to the anonymous membership model. The number and identity of participants in a particular session is unknown, yet, in case of sender charging (and also for inter-provider charging), such values influence the final bill. Hence, one provider would typically charge the other for the required effort to serve a multicast session with a specified number of participants/overall resource consumption. But, due to the anonymity of IP multicast, the effort of one provider cannot be controlled by another provider, the latter just has to trust the former.

This also raises questions about security, trust and related issues which have been mentioned already in Section 4 as prerequisites for any solution. Hence, according security infrastructures and mechanisms must be used for the information exchange among the components.

6 Issues for Business Models

In order to attract and justify the necessary investments into communication infrastructure, it is essential to have a clear business model, which ensures profitable operation of a communication network in a competitive environment. Revenues may be gained from the provision of pure Internet services up to value-added service provision, even including content. But as identified by [Hus99], the pure basic Internet service access provisioning may not show any chances of revenues any more. However, the search for appropriate business models has to be carried out by carefully analysing the relationship of economic and technical aspects of service provisioning. One crucial aspect is given by the potential of product differentiation for network services and value added-services, resulting in strategies which service portfolio is potentially successful. To this end, it is largely impossible to predict future trends in both areas, since the market situation in the current Internet is changing rapidly and prerequisites change almost every week. Therefore, we can only give an overview about the relevant issues.

Technical Aspects of Service Provision

The large range of proposals for providing multi-service communication networks has been presented and classified in Section 2. The notion of network and service performance includes the transmission quality expressed for packets or flows as well as the blocking probability for flows which need guaranteed performance. It might be possible to assemble flow-based guarantees from a service that specifies only packet-based performance objectives, but it is speculation whether this satisfies all eventual applications and whether it comes at the cost of additional complexity at higher layers. Furthermore, it is important to notice the inherent problem of a network specifying flow or packet performance, whereas the user's notion of performance is tied to a complete application session or a transaction, which might consist of multiple network flows or simply a small set of inter-related packets.

Demand Structure

So far, the design of commercial communication networks has been intertwined with a fixed small set of services that are delivered across them. Examples are given by telephony or TV networks. For a limited number of application services, it might be possible to predict price and demand patterns and use this information for capacity engineering. However, the range of applications might be large and diverse in an integrated services network. In that case, prediction of demand and appropriate capacity engineering can be expected to become rather imprecise. The ever-possible advent of new attractive applications might change demand patterns dramatically. For example, a rapid and significant change of demand has been observed in the Internet after introduction of the World Wide Web. Therefore, a stable point of operation might only exist temporarily and as such, has only limited value for capacity planning. In case of rapidly changing demand patterns, network performance as observed by its users might be degraded. However, during such phases of capacity adaptation, service degradation should be controllable and acceptable to users and should not lead to complete service disruption. Therefore, it is important to create a robust infrastructure that allows for controlled service degradation instead of ceasing operation. As mentioned, the level of service degradation has to be assessed by the user's notion of an application session.

Predictability

Depending on the type of application and the particular usage scenario, users' behaviour can be classified along the axes of *price-awareness* and *risk-awareness*. This classification implies the level of price and service predictability that is required. For price-aware users it might be acceptable to cope with fluctuations in prices or differing service performances in terms of packet QoS or flow blocking. In this case, it might also be possible to sell services by auctions. For risk-aware users, predictability of price and service is a crucial objective and the willingness to pay additional charges is high in order to achieve this goal. It is important to notice that similar services are probably used by multiple types of users along this spectrum. As an example, consider a regular private telephone call compared to an emergency call. The users' expectations about the QoS they receive highly influence this classification. There is a possibility for service and price differentiation against competitors by creating different levels of trust in the network's performance.

Concerning the variety of pricing schemes and prices, the main advantage of an auction is determined by the fact that many auctioning algorithms have been shown to have certain desirable economic characteristics [Vic61], besides its price variability. For example, pareto efficient prices (optimal economic efficiency) are achieved during the auctioning process [Var96]. Of course, the stability of a price over time is not achieved, which is contradictory to a certain customer requirement of price transparency as mentioned above. But fixed prices, however, guarantee exactly this transparency in terms of once the pricing scheme will not change, the price for a certain unit service will stay as announced. Nevertheless, this type of price naturally deviates from theoretically optimal prices, because prices have to be based on estimations of price/demand patterns. Therefore, price predictability can be achieved only for a subset of pricing schemes without further adjustments.

Commercial Service Layers

In order to cope with a potentially large variety of demand in terms of service comprehension and predictability, it seems to be an appropriate concept to describe future network operation in terms of different layers of commercial service classes. Such layers may not represent layers in the traditional sense, but they define a logical hierarchy of value-added services, a means for service differentiation and market penetration efforts. Even if there is no technical requirement to create multiple layered service classes, market forces and the providers' adaptation to demand potentially lead to such a situation. Retail entities then provide different interfaces in terms of technical QoS specification and pricing on top of other interfaces. These entities do not necessarily correspond to different institutions. Furthermore, they might not be orderable with respect to their comprehensiveness, i.e., a higher layer does not necessarily represent a richer set of QoS parameters. It is not even clear how many layers are useful and necessary. In Figure 3, we present just one example for an assembly of such commercial service layers. The lowest layer entity sells transmission capacity in large chunks for a fixed a-priori price. Another entity buys this service and creates per-flow QoS service out of it, however, employing varying prices in order to highly utilize the resources. On top of that, a third entity acts as a price insurance broker and resells per-flow QoS service for a fixed price. This service is then used by a video-conference application, probably from a risk-aware user. As an additional example, for certain homogeneous application contexts, it might be possible to built an admission-based service on top of a packet service [GK99a] and trade QoS violation for call blocking. Such a situation can be expressed by the concept of multiple commercial service layers, as well.

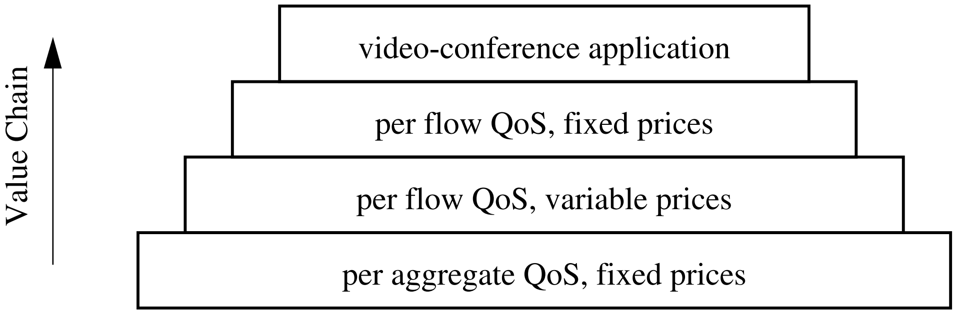

Figure 3: Example for Commercial Service Layers

The tasks of setting prices for services is essential for commercialization. In practice, the highly complex task of marketing determines that setting prices tries to follow a cost-plus basis or it is decided on competitors approaches. Considering the economic theory, these approaches certainly do not define efficient prices. Furthermore, the overall revenue of a provider will be sub-optimal or his presence in a special market segment – the packet-switched networking in the case considered – may be decreasing. Of most interest in a fast growing market, though, the consideration of a variety of influencing factors are required. These factors are not purely technical, such as service quality, access speed, or availability, they encompass a number of market parameters as well, such as service portfolio, service bundles, services types, regional situation, or competitiveness.

Due to the uncertainty about technical developments and demand patterns, the eventual appearance of future networks is unpredictable. Currently, it is particularly unclear which QoS model is superior for the lowest layer of such a layered infrastructure. Different network providers might choose different QoS models and service layers, hence, common interfaces to build useful end-to-end services are of extreme importance.

Vertical Integration

Although the aspects discussed above bear some potential for service differentiation and price discrimination, it remains unclear to which extent service providers might be able to exploit this potential. There is evidence that in a competitive market for network services providing a uniform interface to specify transmission quality, prices will drop to be essentially just cost recovering. In the absence of service differentiation and price discrimination at the network level, it becomes highly important for service providers to approach the integration of transport and application services in order to provide differentiated products. However, this should probably not be considered as the only opportunity. If there is demand for a general network service, this interface should and will be provided. Uncertainty about the aspects discussed in this section and the speed of technological development might create enough space for differentiation.

7 Conclusions

Many issues in the area of charging for packet-switched networks, particularly the Internet, remain open for further research as can be concluded from many uncertain matters addressed in this article. There are mutual interdependencies between the choice of basic network technology for the integration layer, the appropriate business and pricing models, and a charging architecture. Because of literally no real-life experience with providing commercial communication services for a large variety of applications (if anything at all, limited closed user group trials have been performed so far [Var99]), currently, all different proposals are based on plausibility arguments.

Because of the complexity of the overall problem statement, it is extremely important to carry out large-scale simulations and experiments to gather information about the technical applicability of all QoS models and their combinations with service differentiation and pricing schemes. Furthermore, user trials are probably needed in combination with any QoS model, in order to determine demand and payment patterns for such services and their influence on service provisioning. Experiences from the traditional telephone network are certainly important to be considered, but due to the fact that the Internet traffic (at least currently) shows a strong degree of self-similar types of traffic instead of poisson type of traffic [FGHW99], many results from the telephone network need to be re-investigated and cannot be applied simply to the Internet case. Also, it has to be evaluated which user preferences are likely to apply to new service offerings, i.e. whether established assumptions, for example in terms of predictability or payment security, are carried over to such new scenarios or not. In any case, a certain notion of trust among providers, billing services, banks, and users of communication networks has to be established and justified by security mechanisms. In addition, the final architecture of the charging system as proposed needs to be mapped onto the details of the QoS model applied.

To conclude, currently, we do not see any evidence that characterizes any of the above stated choices to be clearly superior to others. Furthermore, it might be very hard if not impossible to create such evidence in experimental setups. This is especially true for scenarios, where technical features are somewhat overlaid by economic considerations, which particularly follow marketing rules or market segment acquisition strategies. Hence, innovative network providers might be forced to simply try out different alternatives for service provision and pricing, and experience their appropriateness in a given market situation.

Acknowledgements

The authors like to thank Bernhard Plattner and the CATI team at TIK, as well as Ralf Steinmetz and the colleagues at KOM for valuable discussions on charging Internet communications.

References

[BBC+98] David L. Black, Steven Blake, Mark A. Carlson, Elwyn Davies, Zheng Wang, and Walter Weiss. An Architecture for Differentiated Services, December 1998. RFC 2475.

[BC99] Joakim Bergkvist and Istvan Cselenyi. Boomerang Protocol Specification, June 1999. Internet Draft, Work in Progress.

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

[BG87] Dimitri Bertsekas and Robert Gallager. *Data Networks*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1987.

[BJS99] Lee Breslau, Sugih Jamin, and Scott Shenker. Measurement-Based Admission Control: What is the Research Agenda? In *Proceedings of the 7th IEEE/IFIP International Workshop on Quality of Service (IWQoS'99), London, UK*, pages 3–5. IEEE/IFIP, June 1999.

[BMR97] Nevil Brownlee, Cyndi Mills, and Greg Ruth. Traffic Flow Measurement: Architecture, January 1997. RFC 2063.

[Bra97] Scott Bradner. Internet Protocol Quality of Service Problem Statement, September 1997. Internet Draft, Work in Progress.

[BZB+97] Robert Braden, Lixia Zhang, Steven Berson, Shai Herzog, and Sugih Jamin. Resource Reservation Protocol (RSVP) - Version 1 Functional Specification, September 1997. RFC 2205.

[CP99] Shaogang Chen and Kihong Park. An Architecture for Noncooperative QoS Provision in Many-Switch Systems. In *Proceedings of IEEE INFOCOM'99*, March 1999.

[Cru91] Rene L. Cruz. A Calculus for Network Delay, Part I: Network Elements in Isolation. *IEEE/ACM Transactions on Information Theory*, 37(1):114–131, January 1991.

[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.

[DB95] Luca Delgrossi and Lou Berger. Internet Stream Protocol Version 2 (ST2) Protocol Specification - Version ST2+, August 1995. RFC 1819.

[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.

[FGHW99] Anja Feldmann, Anna C. Gilbert, Polly Huang, and Walter Willinger. Dynamics of IP traffic: A Study of the Role of Variability and the Impact of Control. In *Proceedings of SIGCOMM'99, Cambridge, MA, USA*, September 1999.

[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.

[Gar95] Roy Gardner. *Games for Business and Economics*. Wiley, New York, USA, 1995.

[GK97] Richard J. Gibbens and Frank P. Kelly. Measurement-Based Connection Admission Control. In *Proceedings of 15th International Teletraffic Congress - ITC 15, Washington, DC, USA*, June 1997.

[GK99a] Richard J. Gibbens and Frank P. Kelly. Distributed Connection Acceptance Control for a Connectionless Network. In *Proceedings of the 16th International Teletraffic Congress - ITC 16, Edinburgh, UK*, June 1999.

[GK99b] Richard J. Gibbens and Frank P. Kelly. Resource Pricing and the Evolution of Congestion Control. *Automatica*, 35, 1999.

[Har68] Garret Hardin. The Tragedy of the Commons. *Science*, 162:1243–1247, 1968.

[HBWW99] Juha Heinanen, Frad Baker, Walter Weiss, and John Wroclawski. Assured Forwarding PHB Group, June 1999. RFC 2597.

[Her99] Shai Herzog. RSVP Extensions for Policy Control, January 1999. Internet Draft, Work in Progress.

[HPW99] Dave Harrington, Randy Presuhn, and Bert Wijnen. An Architecture for Describing SNMP Management Frameworks, April 1999. RFC 2571.

[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.

[Hus99] Geoff Huston. Interconnection, Peering, and Settlements. In *Proceedings of The Global Internet Summit (INET'99), San Jose, California, USA*, June 1999.

[JNP99] Van Jacobson, Kathleen Nichols, and Kedarnath Poduri. An Expedited Forwarding PHB, June 1999. RFC 2598.

[Ker93] Aaron Kershenbaum. *Telecommunications Network Design Algorithms*. McGraw-Hill, 1993.

[KMT98] Frank Kelly, Aman Maulloo, and David Tan. Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[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*. IEEE/IFIP, May 18–20 1998.

[KSWS99] Martin Karsten, Jens Schmitt, Lars Wolf, and Ralf Steinmetz. Provider-Oriented Linear Price Calculation for Integrated Services. In *Proceedings of the 7th IEEE/IFIP International Workshop on Quality of Service (IWQoS'99), London, UK*, pages 174–183. IEEE/IFIP, June 1999.

[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.

[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. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1995.

[MP99] Donal Morris and Verus Pronk. Charging for ATM Services. *IEEE Communications Magazine*, 37(5):133–139, May 1999.

[Odl99] Andrew Odlyzko. Paris Metro Pricing: The minimalist differentiated services solution. In *Proceedings of the 7th IEEE/IFIP International Workshop on Quality of Service (IWQoS'99), London, UK*, pages 159–161. IEEE/IFIP, June 1999.

[PG93] Abhay K. Parekh and Robert G. Gallager. A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case. *IEEE/ACM Transactions on Networking*, 1(3):344–357, June 1993.

[Pos81] Jon Postel. Internet Protocol. Internet Standard, September 1981. RFC 791.

[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.

[RF99] Kadangode K. Ramakrishnan and Sally Floyd. A Proposal to add Explicit Congestion Notification (ECN) to IP, January 1999. RFC 2481.

[RVC99] Eric C. Rosen, Arun Viswanathan, and Ross Callon. Multiprotocol Label Switching Architecture, April 1999. Internet Draft, Work in Progress.

[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.

[SFPW98] Burkhard Stiller, George Fankhauser, Bernhard Plattner, and Nathalie Weiler. Charging and Accounting for Integrated Internet Services - State of the Art, Problems, and Trends. In *Proceedings of The Global Internet Summit (INET'98), Geneva, Switzerland*, July 1998.

[She95a] Scott Shenker. Fundamental Design Issues for the Future Internet. *IEEE Journal on Selected Areas in Communications*, 13(7):1176–1188, September 1995.

[She95b] Scott Shenker. Service Models and Pricing Policies for an Integrated Services Internet. In J. Keller B. Kahin, editor, *'Public Access to the Internet"*, pages 315–337. Prentice Hall, Englewood Cliffs, New Jersey, U.S.A., 1995.

[SK97] David Songhurst and Frank Kelly. Charging Schemes For Multiservice Networks. In *Proceedings of the 15th International Teletraffic Congress - ITC 15, Washington, DC, USA*, June 1997.

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

[Ste97] W. Richard Stevens. TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms, January 1997. RFC 2001.

[SV98] Dimitrios Stiliadis and Anujan Varma. Latency-Rate Servers: A General Model for Analysis of Traffic Scheduling Algorithms. *IEEE/ACM Transactions on Networking*, 6(5):611–624, October 1998.

[Tur86] Jonathan Turner. New Directions in Communications. *IEEE Communications Magazine*, 24(10), October 1986.

[Var96] Hal R. Varian. *Intermediate Microeconomics - A Modern Approach.* W.W.Norton and Company, New York, USA, 1996.

[Var99] Pravin Varaiya. Demand and Provisioning of Quality-Differentiated Internet Access. In *Keynote at IEEE INFOCOM'99*, March 1999. Further information at http://www.INDEX.Berkeley.EDU/.

[Vic61] William Vickrey. Counterspeculation, Auctions and Competetive Sealed Tenders. *The Journal of Finance*, 16:8–37, 1961.

[WLRW99] Richard Waterman, Bill Lahaye, Dan Romascanu, and Steven Waldbusser. Remote Nework Monitoring MIB Extensions for Switched Networks, Version 1.0, June 1999. RFC 2613.

[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.

[Wro97] John Wroclawski. Specification of the Controlled-Load Network Element Service, September 1997. RFC 2211.

[ZSSC97] Zhaohui Zhang, Cheryl Sanchez, Bill Salkewicz, and Eric S. Crawley. Quality of Service Extensions to OSPF, September 1997. Internet Draft, Work in Progress.