An approach to pricing of connectionless network services
Martin Karsten and Ralf Steinmetz
to appear in: *Proceedings of MMNS'97, Chapman & Hall*

An approach to pricing of connectionless network services

Martin Karsten[1] and Ralf Steinmetz[1,2]

1
Industrial Process and System Communications
Dep. of Electrical Engineering and Information Technology
*Darmstadt University of Technology**
Merckstr. 25, 64283 Darmstadt, Germany
phone: +49-6151-16156, fax: +49-6151-166152

2
GMD IPSI
German National Research Center
for Information Technology
Dolivostr. 15 • D-64293 Darmstadt • Germany

{Martin.Karsten,Ralf.Steinmetz}@KOM.th-darmstadt.de

Abstract

In the Internet, the current flat-fee pricing scheme is not suitable to prevent congestion situations. Therefore, pricing is likely to be changed in the future, such that users are charged according to the resources they consume – a usage-based pricing scheme. Because of the heterogeneous network infrastructure, a uniform pricing mechanism must provide for localized pricing policies. It is necessary to allow a step-by-step transition from flat-fee towards usage-based pricing. In this paper, we propose a simple pricing mechanism, *best-effort pricing*, that accomplishes the above goals. Given the connectionless nature of the Internet, we explain why no accurate charging is possible. With best-effort pricing, charges are based on the amount of generated traffic and apply to the sender only. Implications on transport protocols and applications are identified. Finally, we sketch the transition from current flat-fee pricing to best-effort pricing.

Keywords

Network billing, network pricing, connectionless network service, end-to-end communication

---

*. This work is sponsored in part by: Volkswagen-Stiftung, D-30519 Hannover, Germany.

An approach to pricing of connectionless network services
Martin Karsten and Ralf Steinmetz
to appear in: *Proceedings of MMNS'97, Chapman & Hall*

1 INTRODUCTION AND RELATED WORK

Packet- or cell-switched networks form the basis for tomorrows integrated service networks, be it a successor of the Internet or ATM-based B-ISDN (Händel et al. 1994). The most likely scenario is a heterogeneous network infrastructure interconnecting different flavours of subnetwork architectures. Besides the need for connections with a certain guaranteed quality of service (QoS), large parts of data traffic can robustly and efficiently be transmitted using connectionless datagrams. For reasons of clarity, the term *connectionless* is used in this paper to describe data transmission without any state information in routers, also excluding the notion of *soft-states* (Clark 1988). Indeed, both approaches can cooperate, for example using a mechanism like RSVP (Zhang et al. 1993) with IP, where a certain fraction of router and link resources (up to the full capacity) can be reserved, while the remaining capacity is used to handle best-effort traffic.

Pricing will be an issue in these networks. While today large parts of the Internet are used and funded by governmental institutions on a flat-fee basis, we observe the advent of commercial network providers with private and commercial end-users. The provision of integrated services to many end-users (with inherently very different demands for transmission quality), while avoiding the "tragedy of commons" (Hardin 1968, Gupta et al. 1995), will raise the issue of charging users according to their actual use of resources (MacKie-Mason and Varian 1993).

The connectionless Internet architecture has proven to be extremely robust and suitable for data traffic in heterogeneous environments. It is used by millions of users every day. The pricing mechanism proposed in this paper was developed with the current Internet in mind. It is focused on connectionless data transmission to cover existing technology. However, its principles apply to any kind of packet-switched network and can also be used for connection-oriented traffic. This contribution is intended to be a first step towards a pricing mechanism in integrated service networks.

For the purpose of the following discussion, network applications are classified according to two categories, *elastic* and *real-time* (Shenker 1993). Real-time applications need a guaranteed quality of service, for example multimedia applications like video/audio distribution or videoconferencing. A real-time application is expected to use some kind of connection-oriented network service to reserve the needed resources. Subsequently, the network handles the data packets separately from non-reserved traffic to ensure the given QoS guarantees. Elastic applications do not inherently need timing guarantees or a certain bandwidth, i.e. QoS applies only to the correct and orderly delivery of data. Therefore, data are transmitted using connectionless best-effort service, while any QoS demands are satisfied by complimentary mechanisms and protocols (e.g. transport protocol). Examples for such applications are email, file transfer, remote login or web browsing. This paper discusses the implications of pricing on elastic applications. The approach can be seen as an exploitation of the work presented by Shenker et. al. (Shenker et al. 1996). Based on their somewhat general proposal, we present a concrete mechanism and also discuss some implications.

Considerable amount of work on how to calculate prices was done e.g. in (Sairamesh et al. 1995, Wang et al. 1997, Jiang and Jordan 1995). Most of this work tries to derive optimal price calculation formulas. While it is necessary to establish a theoretical foundation for pricing, the practicability of those approaches is contradicted in e.g. (Shenker et al. 1996) and we fully agree with their critique. Additionally, these approaches are applicable to connection-oriented network services only, which does not reflect the currently existing Internet architecture.

The paper is organized as follows: In Section 2, a simple and realistic pricing mechanism is presented. The implications on higher layers up to applications are identified in Section 3, while Section 4 discusses how the transition from flat-fee to usage-based pricing can take place. Finally, Section 5 gives an outlook on future refinements that might be necessary, before usage-based fees can be assessed.

2 BEST EFFORT PRICING

In general, usage-based prices will probably not be the only charges for network connectivity. It is more likely that charges are based on a combination of a connection setup fee, periodical basic fees and usage-based charges (Shenker 1993). This is especially due to the fact that the costs for installing and operating a network are mainly fixed costs, which must be covered independently of the actual utilization. Usage-based prices are intended to regulate the users' generation of traffic as well as to generate revenue to increase the network capacity, if necessary. Thus, usage-based prices are intrinsically related to congestion, while data transmission over non-congested paths (especially in case of best-effort service) can be covered by setup and basic fees. During off-peak hours, it is very likely that communication is essentially free of usage-based charges.

In the following, the term *price* is used to describe usage-based charges. It is important to notice at this point, that pricing does not necessarily refer to a monetary value. Pricing can very well be implemented by quotas, priority levels or a mixture of these (Cocchi et al. 1993).

2.1 Restrictions on pricing of connectionless data traffic

When an elastic application transmits data using a connectionless network service, each packet is treated independently of others. Intermediate systems (routers) are stateless and have very little knowledge about a packet's complete route or inter-packet relations. In such a case, it is almost impossible to guarantee any quality of service or to determine whether the data stream of an application receives a certain service. Packets can be lost, garbled or arbitrarily delayed and there is no instance in the network, which is able to detect such performance degradations. Furthermore, since packets often travel through multiple subnetworks which are operated under different authorities, none of them can solely be held responsible for the orderly delivery of data. For the above reasons, there is a fundamental difference in service charging between the addressed connectionless services, compared to other services, for example telephony: Normally, customers can expect to receive an actual service accomplishment for their expenditures. For connectionless data transmission, it is impossible to guarantee this level of confidence.

In detail, charges cannot be based on the number of successfully transmitted packets only, but must be calculated from the amount of data submitted to the network. That is, a certain well-known price per packet applies for every packet that is submitted to the network, regardless of whether the packet makes it to its destination. Inherently, a single network provider does not have the appropriate knowledge to accurately charge users, hence prices must be based on the amount of data that is *tried* to be transmitted.

For similar reasons, it is rather difficult to charge the receiver of a packet without involvement of additional 'higher level' services. It is clear that a receiver must not be charged without its consent. Therefore, a receiver would have to indicate its willingness to pay for transmission. This can hardly be realized with a connectionless network service.

2.2 General pricing approach

As an important requirement for pricing, it must be possible for users to estimate their costs ahead of time. Therefore, we expect the demand for packet prices (or upper bounds for them) to be well-known for each time, i.e. a pricing scheme is valid during a certain period (~ in the order of days or weeks). A network provider usually has sufficient statistical data to calculate prices for data transmission over its network. As well, the charges that apply for forwarding packets to other providers can be taken into account.

The general approach of best-effort pricing is that a network provider announces its pricing scheme, which is subsequently used by other providers to calculate their prices. In an iterating process, pricing schemes are adapted. However, the aggregated traffic characteristics, especially on major backbone links, are changing slow enough to keep the adaptation periods reasonably long (~ in the order of weeks or months). Hence, prices are well-known and rather static.

2.3 Price calculation

Shenker et al. propose to calculate prices locally, based on the expected route to the target and the expected congestion situation along this route, because in real-world networks it is hardly feasible to derive the corresponding accurate values (Shenker et al. 1996). In accordance with that proposal and opposite to previous optimality approaches, best-effort pricing does not imply a uniform price calculation. As an example, prices could be calculated linearly, depending on the packet size x:

$$p = a + bx + E ,$$

where *a* and *b* reflect the router and link resources, respectively. The parameters depend on the target region (expected route) and current time (expected congestion). The target region determines a subset of outgoing links from the network, so it is sufficient to estimate the effort for transmission within a provider's network domain. *E* is the average price that is expected to be paid to another network provider for forwarding the packet. As mentioned above, a network provider usually has sufficient statistical data to calculate those numbers ahead of time.

The question is not how to actually calculate the packet prices. This must and will be done by every network provider independently. For example, prices do not necessarily always reflect costs, but might very well be subject to marketing considerations. However, it is important to realize that in a connectionless network like the Internet, pricing cannot be done accurately.

2.4 Pricing mechanism

To clarify, we define the following terms: A *service user* is an entity that requests to transmit data over a network domain which is not under its administration. This can be an end-user´s application or a router forwarding packets on behalf of other users. A *service provider* transmits data over its network domain which is neither originated from nor targeted to a host within that domain. The *service interface* is the entity connecting a service user and a service provider. Figure 1 shows the different roles of user and provider along a packet´s route.

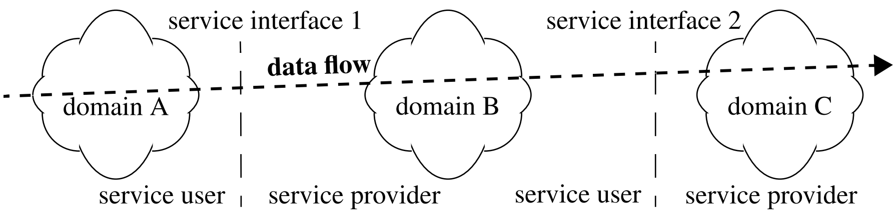

Figure 1 The role of user and provider

Corresponding to the *edge pricing* paradigm (Shenker et al. 1996), charges for a user apply only at the first service interface on the packet´s path. The service provider sets a certain price that applies to every packet the user sends over the provider´s network. This price includes all expenses that subsequently might have to be paid by the provider when the packet is forwarded to another provider. This level of transparency is shown in Figure 2. The user of host α is charged by its provider, regardless of the packet's way through the network. Charges apply only for packets that are accepted by the provider´s network. Hence, if a packet cannot enter a network domain at a certain service interface because of congestion, no charges apply for that packet at this service interface.

Figure 3 shows a complete transmission scenario. When a packet is sent from host α to host β, domain X is a service user at service interface 1. At service inter-

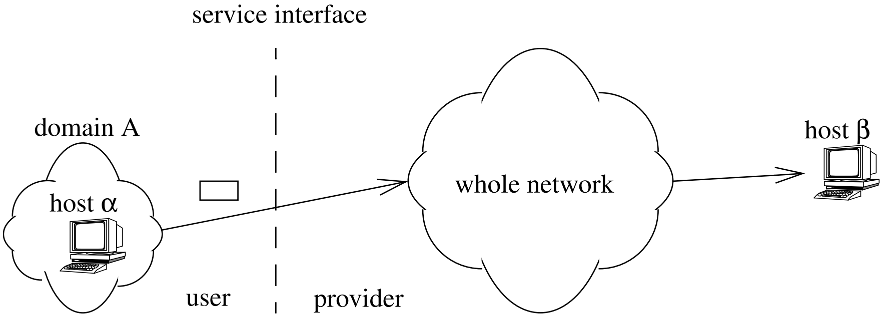

Figure 2 Locality

face 2, domain A is the service user, while domain B is the service provider. Finally, at service interface 3 no charges apply, since the packet is destined to host β in domain Y. In this example, the operator of domain A has announced a price, based on the expected route, for delivering a packet from service interface 1 to β. This price is paid by X. The operator of domain B has announced a price, based on the expected route, for delivering a packet from service interface 2 to β. This price is paid by A. Of course, B´s price from service interface 2 to Y (and prices for alternative routes) influences A's price from service interface 1 to Y.

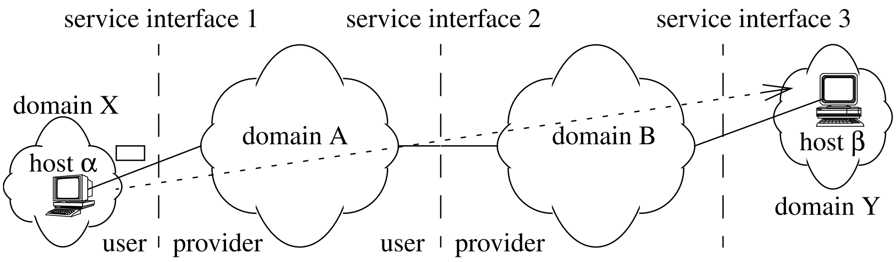

Figure 3 Complete packet transmission

In a second step, the pricing mechanism can be augmented by multiple priority levels, in a way that packets with a higher priority level are subject to higher charges, but are served ahead of lower priority packets, thus receiving potentially better service. Priorities do not necessarily have to be uniformly defined, but they are useful, only if some sort of priority mechanism is implemented in every router along the network path (at least in any router that is congested or that transmits packets over a congested link). Nevertheless, high priorities do not imply any real guarantees.

In order to implement the described pricing scheme, a provider must gather information about the amount of data and/or number of packets transmitted for each quadruple:

`<user, target region, time period, priority>`.

Instead of storing the number of packets for each quadruple, the granularity can be lowered by immediately assigning a price to each packet while it is handled by the router. In this case only the cumulative billing amount has to be stored for each user. Given the fact that routers already collect very extensive information about the processed traffic, this overhead seems to be feasible under today's technological limitations.

If a network architecture provides a connection-oriented service with guaranteed QoS, the same pricing mechanism applies in a way that charges are based on the amount of sent data. Normally, the connection is expected to fulfil its service guarantees, so in this case the amount of sent data reflects the actual usage of resources. However, it is not clear how to handle the case when any provider along the path does not fulfil its part of the QoS contract.

3 IMPLICATIONS

At a first glance, it seems not appropriate to charge network users for only trying to transmit their packets. However, as discussed in the previous section, this is the only way billing can be done at all. Therefore, it is necessary to reason about implications of such pricing. Additionally, further refinements are possible to balance out the inaccuracy of best-effort pricing.

3.1 Applications

Using best-effort pricing, the sender is charged for transmission of data. This simple approach does not fit with all applications of data communication. Consider for example an anonymous ftp server. When a client downloads a file, data flows from server to client, but the client is the one that is mostly interested in this service. In such a case, it would be more appropriate to charge the client for transmission.

An important aspect of pricing data traffic is whether the transmission yields profit for the sender or the receiver. This can be reviewed from a more abstract level. In almost all network applications, data flows in both directions. Nevertheless, the amount of data for each direction can differ substantially. For that reason, it is useful to view the communicating entities as *initiator* or *responder*. Then the most popular network applications can be classified according to the amount of data sent and received by each communication entity. In Table 1, the **relative** share of traffic among initiator and responder is listed. The terms *defacto sender* and *defacto receiver* are used from now on to describe the distribution of traffic generation. At first, only the simple cases are considered in which the initiator is interested in communication.

It turns out that for the purpose of pricing, communication scenarios can be separated into two categories, depending on the relation of the initiator to the responder. In *anonymous* scenarios, initiator and responder do not necessarily have any relation and the vast amount of traffic flows from responder to initiator. In *registered* scenar-

Table 1 Relative fraction of traffic for typical network applications

	email	remote login	normal ftp (put / get)	distr. applications (e.g. shared whiteboard)	web browsing	anon. ftp
traffic from initiator	high	medium	high / low	medium	low	low
traffic from responder	low	medium	low / high	medium	high	high
category			registered		anonymous	

ios, the initiator is well-known to the responder and the distribution of sent and received traffic can be arbitrary.

If the initiator is the defacto sender as it is the case in some registered scenarios, it is clearly suitable to charge him for data transmission. On the other hand, if the responder is the defacto sender or both parties send considerable amounts of data, charges apply to the responder, although the initiator is primarily interested in communication. However, the initiator is well-known to the responder, so charges can be passed on to him, according to the responder´s policy. Hence, for the purpose of network service pricing, it is appropriate to charge the responder. Given these considerations, it is acceptable to always bill the sender in registered scenarios.

In anonymous scenarios, the defacto sender (responder) has little benefit from the data transfer and can hardly control the amount of data that is transmitted. Therefore, the defacto sender should not be charged for data transmission. Since it is impossible to directly charge the receiver, We suggest to use two basic priority levels to deal with anonymous scenarios. Priority level 0 indicates an anonymous scenario, while priority level 1 indicates a registered scenario. Packets with priority 1 are always served ahead of packets with priority 0. Packets with priority 0 are priced much less, maybe nothing, and only use the capacity that is not used for level 1 packets. Charges for priority 0 packets can be embedded in the periodical basic fee.

On the other hand, the responder in anonymous scenarios might be interested in sending data with better service. Examples are commercial web pages, for which access is charged anyway. In this case the provider could choose to send the data with priority 1 and include the charges into the normal access price. As well, a company might consider the public relation effect of its web presentation very valuable. Then, the company can send the data traffic with priority 1 and consider the charges as marketing costs.

3.2 Refinements

Network providers can precisely calculate their prices based on traffic and congestion estimations. If any network provider tries to 'cheat', the market mechanism will bear another one with fair prices. Users can generate traffic statistics to monitor the billing process and to optimize their generated network load. If a user encounters significant performance degradation, in the short run, she can decide to either postpone the request or to increase the priority level. In this case, since the price for a few packets will be negligibly low (MacKie-Mason and Varian 1994), an automated adaptation algorithms can quickly determine the desired reaction (see Subsection 3.3). In the long run, users might consider to switch to another network provider, if they receive unsatisfying service.

On the other hand, a provider can decide to consider prices as upper bound values. In case an outgoing link from the provider´s network is congested and does not accept as much traffic as expected for a certain period of time, the provider's charges for this link are decreased and the provider can give refunds to the users who tried to send data across the link. Those refunds are statistically split among the users according to the amount of data that was tried to sent. The recursive application of such a refund mechanism up to the sending user balances out the inherent inaccuracy of best-effort pricing.

For example, the German backbone for academic and research institutions is run by the *DFN-Verein* (DFN 1997), which is a non-profit organization. If a single network link is split up within a large institution, a networking division operating the internal network can be seen as a non-profit network provider, as well. In such situations, the primary non-profit provider can easily set internal prices that are high enough to ensure sufficient revenue for the network infrastructure and to pay external service provider(s). The above refund mechanism can be used to increase fairness among service users of the non-profit network provider.

3.3 End-to-End communication

There are several problems that may occur to a packet on its way through the network. While elastic applications do not inherently rely on certain timing or bandwidth guarantees, they usually cannot tolerate lost or garbled packets. Using retransmission to recover from lost or garbled packets, all kinds of errors result in increased delay. For elastic applications, the average delay of packet transmission largely determines the level of quality as it is perceived by the user. Hence, when using retransmission in a best-effort priced network, lost or garbled packets cause even two unpleasant effects:

- The transmission becomes more expensive, since charging is based on the amount of sent data.
- Quality is degraded because the average end-to-end delay is increased.

Another possibility to guarantee correct delivery of data is the use of *forward error correction (FEC)* to deal with lost and garbled packets. Using FEC avoids increases in delay, but imposes additional transmission costs, because the amount of sent data is permanently increased.

In general, it is necessary for transport protocols to quickly determine congestion and to react accordingly. For example, when a TCP entity encounters congestion between end-systems, it uses the *slow start* algorithm in conjunction with a *treshold* mechanism to quickly decrease the data rate (Tanenbaum 1996). It then increases the sending rate, first quickly, later on slower, to determine which rate is appropriate. This mechanism perfectly fits with best-effort pricing, because it avoids sending even more packets that would have to be paid, but very probably are lost during transmission.

3.4 More aspects on applications

The discussion above left out a certain class of applications, which are quite popular, as well. Examples for those services are news and mailing lists. These services store and forward data on behalf of third parties, but the defacto sender also has a certain interest in the distribution. To this end, the best solution seems to shift the transmission of these data to off-peak hours, as it is currently done for news at a lot of sites.

Another issue are proxy and cache servers. With best-effort pricing, their network domain is charged for transmission of data. This does not constitute a problem. Mostly, such a server resides within the same domain as its users. If not, then clients are probably charged for accessing the server, anyway. Hence, the traffic charges can be embedded into access charges. Another option is for the server to keep track of the amount of data that is sent to a certain client and pass on the charges.

4 TRANSITION ISSUES

To set up a pricing mechanism in the Internet seems to be rather difficult. A price calculation method that tries to achieve optimality must be implemented by every network provider. A price mechanism that relies on a certain packet format or a special protocol must undergo the rather lengthy standardization process. It then has to be deployed into the network requiring changes to all sites. This is an unrealistic assumption for the near future. Both aspects are not a problem for best-effort pricing. A single provider can estimate the cost of forwarding packets over outgoing links by dividing the flat-fee for that link by the amount of data that is usually sent across that link. If any outgoing link is priced using best-effort pricing, a provider can use the announced pricing scheme to calculate its own prices. Two priority classes for anonymous and registered application scenarios must only be supported by providers that support best-effort pricing in the first place. In IPv6 (Deering and Hinden 1995), a new options header can be defined while in IPv4 a new header option is needed to support such priorities (or maybe the TOS field can be used). Given the fact that only a single bit is necessary to implement the minimum needed two priority levels, a more efficient solution might be agreed upon, e.g. using a reserved bit. Until those issues are resolved, providers can develop a workaround (e.g. by using a reserved bit without standardization).

Given the above considerations, it is easily possible for any network provider to implement best-effort pricing regardless of other providers´ opinion on this issue. Therefore, a seamless transition towards usage-based pricing seems to be feasible.

5 SUMMARY AND FUTURE WORK

As shown in this contribution, pricing in connectionless networks cannot be accurate. Considering this, a practical pricing mechanism, best-effort pricing, is proposed that is based on the amount of sent data. In correspondence with previous research results, best-effort pricing provides localized control of pricing and billing. To balance out the inherent inaccuracy, a refund mechanism is suggested. Implications on transport protocols are discussed. To deal with data traffic that is generated for the receiver´s benefit, applications are classified and the need for at least 2 basic priority levels is identified. The demanded transition from flat-fee to best-effort pricing is presented. In conclusion, the feasibility to actually implement best-effort pricing is shown by proof of concept.

However, considerable future work is needed to fully cover all details of pricing in packet- or cell-switched networks. First of all, further simulations and implementations are needed to back up the proposed implications of best-effort pricing. Calculations based on real world traffic statistics might be a first step in this direction. As well, more work is needed to derive good approximations for the optimal setting of prices for connections with a guaranteed QoS. To this end, it is not understood how to reasonably charge for multicast communication. Providing of multicast and QoS is an active research area, so it might be helpful to address the problem of pricing together with the development of such services.

ACKNOWLEDGEMENTS

We would like to thank Lars Wolf and Jens Schmitt for reading draft versions of this paper and providing valuable comments.

REFERENCES

Clark, D. (1988). The Design Philosophy of the DARPA Internet Protocols. In *Proceedings of ACM SIGCOMM '88*.

Cocchi, R., Shenker, S., Estrin, D., and Zhang, L. (1993). Pricing in Computer Networks: Motivation, Formulation, and Example. *ACM/IEEE Transactions on Networking*, 1(6):614–627.

Deering, S. and Hinden, R. (1995). RFC 1883 - Internet Protocol, Version 6 (IPv6) Specification. Internet Draft.

DFN (1997). WWW-Server des DFN-Vereins - Homepage. http://www.dfn.de/.

Gupta, A., Stahl, D. O., and Whinston, A. B. (1995). The Internet: A Future Tragedy of the Commons? In *Conference on Interoperability and the Economics of Information Infrastructure*. avaiable from http://cism.bus.utexas.edu/alok/wash_pap/wash_pap.html.

Händel, R., Huber, M. N., and Schröder, S. (1994). *ATM networks*. Addison-Wesley, 2nd edition.

Hardin, G. (1968). The Tragedy of the Commons. *Science*, 162:1243–1247. available from http://www.csra.net/LRAnd/GH_tragc.htm.

Jiang, H. and Jordan, S. (1995). The Role of Pricing in the Connection Establishment Process. *European Transactions on Telecommunications*, 6(4):421–429.

MacKie-Mason, J. K. and Varian, H. R. (1993). Pricing the Internet. In *"Public Access to the Internet"*. JFK School of Government, Harvard University. available from http://www.spp.umich.edu/spp/papers/jmm/Pricing_the_Internet.ps.Z.

MacKie-Mason, J. K. and Varian, H. R. (1994). Some FAQs about Usage-Based Pricing. available from http://www.spp.umich.edu/ipps/papers/info-nets/useFAQs/useFAQs.html.

Sairamesh, J., Ferguson, D. F., and Yemini, Y. (1995). 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, Los Alamitos, CA, USA. IEEE Computer Society Press.

Shenker, S. (1993). Service Models and Pricing Policies for an Integrated Services Internet. In *"Public Access to the Internet"*. JFK School of Government, Harvard University. available from ftp://parcftp.xerox.com/pub/net-research/policy.ps.Z.

Shenker, S., Clark, D., Estrin, D., and Herzog, S. (1996). Pricing in Computer Networks: Reshaping the Research Agenda. *ACM Computer Communication Review*, 26(2):19–43.

Tanenbaum, A. S. (1996). *Computer Networks*, chapter 6.4.6, pages 536–539. Prentice Hall, 3rd edition.

Wang, Q., Peha, J. M., and Sirbu, M. A. (to appear in 1997). Optimal Pricing for Integrated-Services Networks with Guaranteed Quality of Service. In Bailey, J. and McKnight, L., editors, *Internet Economics*. MIT Press. available from http://www.ece.cmu.edu/afs/ece/usr/peha/peha.html.

Zhang, L., Deering, S., Estrin, D., Shenker, S., and Zappala, D. (1993). RSVP: A New Resource ReSerVation Protocol. *IEEE Network Magazine*, 7(5):8–18.