Charging for ATM-based IP Multicast Services

Georg Carle
GMD FOKUS
Kaiserin-Augusta-Allee 31
Berlin, Germany
Tel.: +49 30 3463-7149
Fax: +49 30 3463-8000
E-Mail: carle@fokus.gmd.de

George Fankhauser, Burkhard Stiller
Computer Engineering and Networks Laboratory, TIK
Swiss Federal Institute of Technology, ETH Zürich
Switzerland
Phone: +41 1 632 [7017 | 7016]
Fax: +41 1 632 1035
E-Mail: [gfa | stiller] @ tik.ee.ethz.ch

Abstract

Charging and accounting of Internet services has been addressed in numerous proposals. However, finding a solution suitable for ATM-based IP multicast services is highly challenging. IP multicast over ATM introduces high complexity, in particular for multicast with heterogeneous QoS and separate administrative IP domains by involvement of multiple Internet service providers. This paper proposes a framework for charging of IP multicast over ATM that is suitable for heterogeneous scenarios with multiple interconnected IP service providers and non-negligible losses. A scheme for ATM-based IP multicast transmission of wavelet-encoded video is presented that uses multiple ATM service classes. The scheme supports multicast scenarios with heterogeneous QoS and allows to trade off video quality, error tolerance, end-to-end delay and costs.

1. Charging and Accounting

Charging and accounting of IP services gain more and more interest as the Internet is moving from an academic to a commercially dominated environment. A satisfying charging model has to deal with the following three main aspects of charging:

1. *Customer Aspects*: The customer has to understand the charging model and must be aware of the costs he produces when using a service. The costs should not differ significantly from costs of the telephone tariffing system the customer is used to. This introduces a threshold of acceptance. Costs should be based on goodput, i.e. errors within the network that require additional traffic for error recovery should not lead to increased costs for the customer.
2. *Economic Aspects*: The model has to differentiate between several types of service and – within these types – between different levels of service quality. Thus, the customer is animated to choose the cheapest type of service to satisfy his needs. The costs can be assigned to the customer.
3. *Technical Aspects*: In order to minimize the effort for computing the different tariffs the charging scheme has to be as simple as possible. After having distinguished between different types of services the charging model can be optimized for each type. As cell-based charging in ATM networks is computationally demanding, dedicated hardware-support may be required.

To summarize these three aspects, charging of ATM-based IP multicast services should be kept sufficiently simple to be feasible and intelligible, while allowing to differentiate between different types of service.

State of the Art

First attempts to introduce usage-based pricing and related charging mechanisms aimed at congestion avoidance and tried to encourage efficient use of network resources. Furthermore, the differentiated pricing of different quality of service (QoS) levels revealed a gain in utilization of the network for heterogeneous use (i.e. for a variety of applications and usage patterns). Cocchi et al. [CoES93] introduced a simple priority model with an exemplary price discrimination for different levels of service giving preference to bandwidth and/or delay sensitive applications. Their simulation results proof that flat pricing is inferior to a priority model. However, the assumption of fixed base prices for service classes and the difficulty of describing user utility functions leave the applicability questionable. MacKie-Mason and Varian [MaVa95] applied basic economic theory to include network externalities in the user's price for a congested resource, typically the queue of an outgoing link in a router. To implement this pricing scheme they introduced a new clearing method based on auctions [MaVa94] which aims at obtaining real-time market prices. They argue that in sufficiently competitive markets all QoS levels will be priced efficiently. In [Shen95] these ideas are refined by studying the utility/bandwidth curves for different applications. Architectural design issues for the future Internet are discussed, advocating multiple service classes in combination with admission control and usage-based pricing.

In [EdMV95] an approach based on flow measurements of TCP traffic is presented. The performance problem associated with measuring traffic is solved by using modified router lookup hardware. In addition, authentication, price feedback to users and payment control are discussed. Based on a modified TCP, Clark and Fang propose in [ClFa97] an approach called Expected Capacity Framework which does not only provide usage sensitive pricing but also introduces soft QoS guarantees. Recently, proposals for charging of reservation-based systems, such as RSVP, emerged. In [MacK97], the auction-based "Smart Market" approach is re-examined to be used with reserved flows. [FSVP98] presents an example implementation of such a pricing and charging system for reserved flow in an RSVP-based Internet.

The problems discussed in related work fall into two categories: Price determination and structural issues. Charging and accounting overhead can influence pricing schemes. Charging protocols may prevent the use of provider-specific pricing. Pricing depends on the mechanisms and systems used, because the infrastructure imposes limits on the granularity of the priced goods (*e.g.*, cell, packet or flow based methods). However, current pricing schemes and proposals for integrated services reflect a wide range of opinions and arguments.

- The goal of pricing is controversial. Prices can be used for users to allow most efficient resource sharing or for a provider to maximize revenue (e.g., see [RaRK96]). For the common goal of efficient resource sharing it is not clear how gained efficiency is to be distributed. Pricing of congestible resources that yield optimal prices for users (at least theoretically) may not allow service providers to fully recover their cost [ShCE96].
- Some pricing models only observe local cases (single nodes). For a full network global optimality and stability are important, but not necessarily needed. [ShCE96] advocates ”Edge Pricing”, where prices can be determined locally rather than computed in a distributed fashion along the entire path. This is very attractive, as it limits the complexity of the pricing scheme and as it provides a high degree of independence for accounting schemes of different service providers.
- Multiple provider along paths usually do not want to implement the same pricing models, because of competition and differing business culture. This leads to different accounting domains and dynamic peering contracts. This scenario is discussed in [ClFa97], but open issues such as highly changing QoS requests remain.
- Different services classes may require different pricing. ATM tariffs may differ largely for the different ATM service classes that can be used for providing IP services. The role of distance and duration of a connection may decease. One example is the proposed set of ATM tariffs for VBR, CBR, ABR, and UBR by Walker et. al. (see [WaKS97]), where the price per volume for VBR traffic is two orders of magnitude higher than the price per volume for UBR service. These prices are set independently of a distance and duration to focus on the error-rate obtained, only. This price difference gives a high incentive to select the UBR service class whenever possible. However, a pure volume-based approach for all service classes as in [WaKS97] does not reflect the complete picture. Since constant bit rate service classes are in principle very similar to the traditional telephone system, a distance-related component may be necessary to incorporate the provider perspective.

Architectural Considerations on Charging and Accounting

Measurements are suitable for providing adequate input to fair charging and accounting methods [CoKW97]. However, overhead of fine grained measurement-based charging approaches are a major and important concern. Proposed solutions include moving the measurement to the edge of the network [ShCE96], integrating specialized hardware or software modules into switches and routers [EdMV95, CaSS96a, CaSS96b, DWDA97], dedicated measurement devices and statistical sampling methods. Contractual traffic estimation based on user profiles [ClFa97] or explicit reservations [FaSP98] allow to avoid measurements. Results in implementing two pricing schemes show that a low overhead for the reservation-based approach can be achieved [FaSP98].

Cost Sharing between Receivers and Senders

Multicasting introduces the problem of *cost sharing* between different receivers of a multicast tree [HeSE97]. Herzog [Herz96] presents a cost sharing approach for multicast trees. However, balancing of the multicast tree for a highly dynamic sender and receiver behavior remains to be investigated.

Shenker et al. [ShCE96] present a critical view on pricing in computer networks. They show that it is attractive to support receiver-based payment or a sharing of payment at the network level. One cases for cost sharing is information retrieval of news data from a multimedia database, where traffic is asymmetric. Web-based applications also suggests a receiver-based payment.

It can be debated whether the functionality for indicating the willingness of a receiver to pay for the connection should be provided at the network layer. It will always be possible to use only sender-pay communication at the network layer, and to adopt a higher-layer protocol to reimburse the sender. However, this approach is difficult when multiple service providers are involved, and when receivers want to share fairly the costs related to individual QoS properties.

An explicit protocol can be used to signal whether and to which extend a receiver is willing to pay. For a path that traverses several service providers, this information needs to traverse the path. An alternative scheme for allowing payment to be performed by either sender or receiver that does not need explicit signalling is to adopt a uniform standard, such as reserving a certain portion of the address space for receiver-pay groups.

2. Reliability and Error Control

The provision of Internet Integrated Services using ATM technology allows to provide highly reliable Internet services. However, economic motivation for achieving a high statistical multiplexing gain is a main driving force for operating the network in a way that non-negligible error rates occur.

For the provisioning of reliable services, error control mechanisms are required that detect and correct corrupted and lost packets or cells. Depending on the origin of the errors, different error characteristics may be observed. As a consequence, different error control mechanisms may be identified to be more appropriate in a certain case.

In order to meet the reliability requirements of a real-time application, it is possible to use a network service that directly provides the required reliability, without additional error control mechanisms in the transport layer. This can be ensured by reserving network resources for the application, or by dimensioning the network in a way that the residual error probability is sufficiently small (over-engineering).

When using network services that do not guarantee the Quality of Service (QoS) required by an audio-visual application, error control schemes are required for recovery from losses due to congestion in the network. Existing transport-layer error control mechanisms with Automatic Repeat Request (ARQ) and Forward Error Correction (FEC) are presented in [CaBi97], including a discussion of their suitability for use in IP-based networks.

The appropriate combination of ARQ and FEC is the key to achieve optimal performance. Hybrid ARQ where initial transmission is protected by redundant information provides improved delay characteristics. Retransmission of parity allows for scalability for large receiver groups and most efficient usage of bandwidth. By providing sufficient transport protocol processing capability with a low latency, it is possible to meet delay requirements of many audio-visual applications even after one or two retransmissions. This strategy potentially offers better utilization of network resources in particular for highly bursty sources.

Error Control for Layered Video Services

Compressed video streams have many challenging properties: high data rates in comparison with many other internet applications, high reliability requirements, complex interdependencies between different packets and varying delay requirements.

A challenging property of video streams is their relatively high data rate (e.g. 1.5 Mbit/s for MPEG-1). Such a high bit rate poses the problem of rate control to multiple receivers: One way to allow for rate control and scalability is to use hierarchical coding schemes, where the signal is encoded in a base layer that provides a low quality image and additional complementary layers for improved image quality. Each receiver needs to receive at least the base layer. One of these approaches has been implemented in a WaveVideo approach mapped onto ATM [DaFP96]. Different layers of the video signal are tagged and transmitted on a single multicast group. Receivers choose how many layers in addition to the base layer they want to receive by signaling these requests to network filters that operate on the tagged video stream. Another approach, referred to as receiver-driven layered multicast or RLM for short uses different multicast groups to represent video layers [McVJ97]. Li, Paul, Pancha and Ammar [LiPP97] have recently proposed a retransmission-based loss recovery protocol, called Layered Video Multicast with Retransmission (LVMR), for non-interactive transmission of MPEG video to multiple receivers across the Internet. The experiments performed for a playout delay of at least 1500 msec indicate that LVMR is able to recover around 80% of the losses seen by a receiver.

Charging and Error Control

High-quality ATM streams allow for provision of IP services with a service quality directly sufficient to provide high-quality voice and video services. Cheap UBR ATM streams result in IP services that do not always meet application requirements directly. However, by providing transport-layer error recovery mechanism with FEC and/or ARQ, it is frequently possible to provide audio-visual services with significantly lower costs that still meet user requirements.

Recognizing error control capabilities of end-to-end error control schemes and regarding price differences for Integrated Internet Services makes the following strategy attractive: Using a very cheap basic network layer and using additional layers for increasing reliability will reduce the costs for transmissions. As losses mainly originate from congestion, redundancy or retransmissions transmitted over the same service category as the packets to be protected may suffer from high losses at times of congestion. If forward error correction is applied to the original data, the service quality has to be distinguished within one flow of data. Appropriate coding schemes allow for a separation between data representing the original data with little overhead.

3. Layered Video Transmission over ATM-based IP Multicast Services

Considering the service provision value chain with a network operator, a value added service provider (VASP) and Internet users as described in [BoCC97], we regard the example of providing streaming video to Internet users.

It is assumed that the service is interactive (i.e., low end-to-end delay budget) and the video applications use a coding scheme which is able to exploit multiple qualities of service. Such video coding schemes were developed to provide receiver controlled quality selection in multicast scenarios (Progressive Video with Hybrid transform, PVH, [McVJ97]) and error-tolerance in environments exhibiting significant loss rates (c.f. [HoSp96] and WaveVideo [DaFP96]).

Most of these coding schemes are based on wavelet transforms. For transmission, portions of the coded video stream having different properties (spatial, color, and temporal resolution) are either sent on separate connections (c.f. Figure 1a) or tagged and sent on a single connection (c.f. Figure 1b). In the first approach receivers select a certain level of quality by adding up layers of non-redundantly coded video streams. The second approach employs only a single connection and tags packets according to their contents. Receivers can then subtract undesired properties from the stream.

For the example, both of these schemes can be applied. The cost optimization is based on the fact that certain portions of the video are more important than others. The low-frequency subband (decoded by D1 in Figure 1) forms the root of the image hierarchy and is most vulnerable to cell or packet loss. Using IP services with poor reliability, as in the case of UBR may therefore lead to significant video quality degradations. Using transport services that are not capable of providing different service levels, the critical part of the video stream needs further protection by additional transmission of redundancy, or by using an ARQ protocol which would add additional delay.

In an ATM environment the problem of poor video quality can be resolved by using different ATM service classes for low- and high-frequency components of the video stream. Since loss of high-frequency components has less impact on video quality than loss of low-frequency components, ATM UBR services are frequently suitable for their transmission. A desired minimal quality can be assured by transmitting low-frequency components over ATM service categories with guaranteed service quality, such as VBR or CBR.

Using a pricing model for UBR and VBR services, as described in [WaKS97], the cost c for transmission of the low- and high-frequency streams can be given as the sum of the low- and high-frequency volumes multiplied by the prices for VBR and UBR traffic classes. Varying the aspect ratio of low- and high-frequency volumes at the VASP (we assume that the Internet user wants to limit the total cost for a single video stream), the total cost for this connection paid

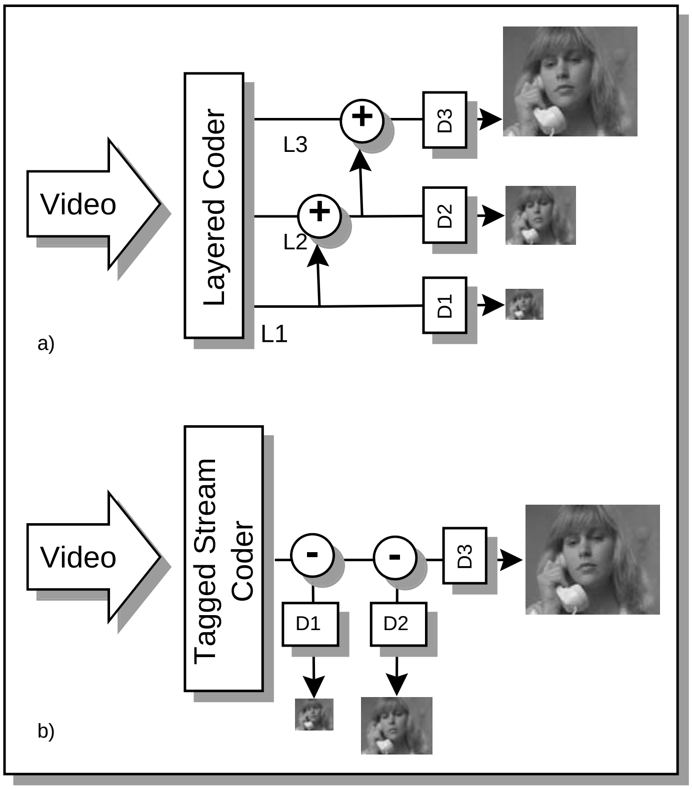

Figure 1: Layered video coding with (a) separate streams or (b) single tagged video stream

to the network operator is plotted by the dashed curve in Figure 2. According to the aspect ratio which runs from low-frequency (showing a small coarse-grained picture only) to high-frequency only, the dotted line gives the qualitative behavior of the user's utility u approximated by average video quality which can be obtained by a technical measure (*e.g.*, an approximation of the peak signal-to-noise ratio by the number of received packets) or a poll among users (*e.g.*, mean opinion score). In order to compare this utility with the resulting cost along the Low/High-ratio-axis (L/H), it must be scaled to yield a monetary unit. This factor includes the valuation of this service by the Internet user. Using the difference of utility and cost for maximization gives VASPs an optimal choice for application configuration (L/H ratio) for a given price set of UBR and VBR services.

If this simple scenario is to be extended to a more general setup which includes multicast communications and multiple network operators (in the figure abbreviated as “no”) in the same service area, the optimal configuration of the video application has no longer a simple solution. Consider a network configuration as in Figure 3. It is a desirable property in large networks that network operators may choose their prices locally to their accounting domains [ShCE96]. Starting from the leaves of the delivery tree towards the server, the above optimization can be applied based on VBR and UBR prices of network operator no2 and no3. Using the utility curve of the application, the optimal VBR to UBR ratio can be calculated for each accounting domain. With differing prices these ratios will also differ. Now there are

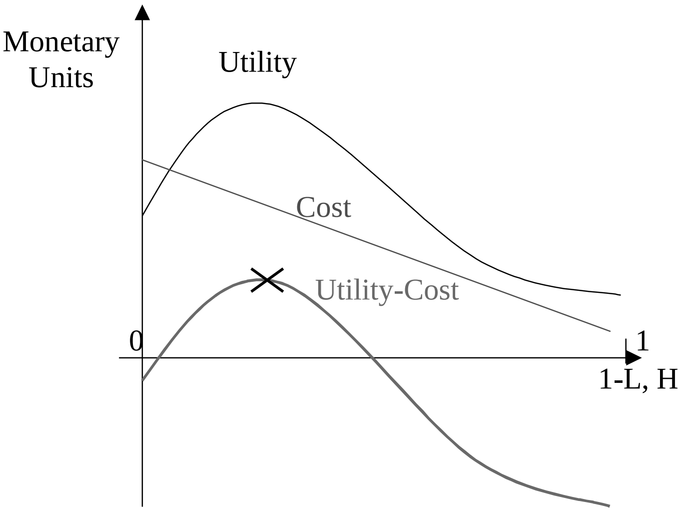

Figure 2: Cost and utility functions for layered video

two possible strategies for the VASP to configure the video service:

- To run all subtrees at local optima, the bandwidth requirements for VBR and UBR must be maximized where providers are interconnected. At switch s2 the bandwidth requirements for no2 and no3 are maximized and these new values are then propagated to the server to include no1 in the process. This method will lead to higher bandwidth allocation and costs towards the root of multicast trees. Since this growth of bandwidth depends only on the variation of the pricing schemes of the network operators which should converge in an efficient market, the method of maximizing bandwidths can be considered as simple and feasible.
- Another approach to find a global solution (e.g. if price variation among network operators is high) is to average the bandwidth ratio at each point where a new network operator joins the tree. The weight of each subtree depends on its number of clients and the price. This method is again applied from leaves to the root. Whenever a new ratio is averaged it is propagated upwards with the total price-bandwidth product for the joined subtree as a new weight.

4. Conclusions

This paper describes the aspects of service allocation in a networking environment offering price discrimination for ATM service classes. The discussion encompasses applications running on UBR services with non-negligible cell and packet loss, IP multicast services over ATM with support for reliable transmission, value added IP multicast services and multiple service providers.

Using advanced applications, such as layered media coders, it is possible to obtain better service or lower cost by splitting the media stream in a loss-sensitive and a loss-tolerant substream than by using a single service class which is used to transmit the whole media stream.

Using the described error-control schemes, costs can be reduced significantly. Simple applications using single

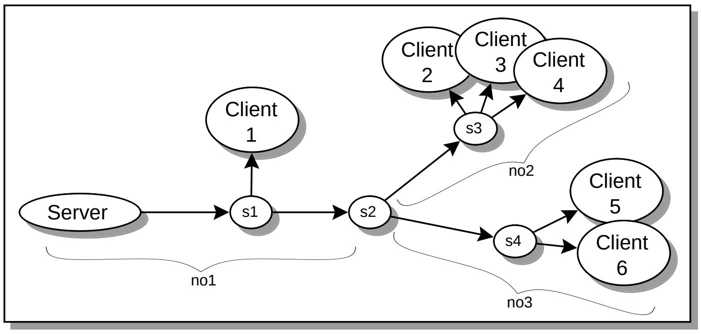

Figure 3: Multicast with multiple service providers

streams for network transport are configured to use the low-cost ATM service UBR. In order to eliminate bursty loss caused by congestion, an error control channel is introduced using more expensive service class with higher reliability such as VBR. Due to the bursty nature of cell losses and the implication on framing (loss of whole AAL 5 frames or IP packets when single cells are lost), a large potential gain from FEC and ARQ schemes on a separate channel with a higher quality of service can be achieved. In particular for large groups this scheme has a significant impact on scalable low-cost services for continuous-media applications.

Future work in the field of service and cost allocation includes simulations and quantification of the methods described in this paper. Besides various application types (error tolerant, explicit error-control) and new cost-sensitive transport protocols, the implications of pricing schemes need to be studied, and quantitative results for parameterization need to be obtained.

References

[BoCC97] D. Botvich, Y. Chen, T. Curran, B. Kerswell, J. McGibney, D. Morris: On Charging for Internet Services Provided over an ATM Network, IEEE Workshop ATM'97, Lisbon, Portugal, May 1997.

[CaBi97] G. Carle, E. W. Biersack: Survey on Error Recovery for IP-based Audio-Visual Multicast Applications, IEEE Network Magazine, November/December 1997, pp. 2-14.

[CaSS96a] G. Carle, G. Schäfer, J. Schiller: An Approach to Hardware-Supported Accounting Management in ATM-Networks, 8th IEEE Workshop on Local and Metropolitan Area Networks, Berlin/Potsdam, Germany, August 1996.

[CaSS96b] G. Carle, T. Schmidt, J. Seitz: Fair ATM Charging with Consideration of Traffic Characteristics and QoS Parameters, IEE Colloquium on "Charging for ATM", London, U.K., November 1996.

[CoES93] R. Cocchi, D. Estrin, S. Shenker, L. Zhang: Pricing in Computer Networks: Motivation, Formulation, and Example; IEEE/ACM Transactions on Networking, Vol. 1, No. 6, December 1993, pp. 614-627.

[ClFa97] D. D. Clark, W. Fang: Explicit Allocation of Best Effort Packet Delivery Service, Technical Report, Laboratory for Computer Sciences, MIT, 1997.

[CoKW97] C. Courcoubetis, F. Kelly, R. Weber: Measurement-based charging in communication networks, Statistical Laboratory Research Report 1997-19, University of Cambridge, 1997.

[DaFP96] M. Dasen, G. Fankhauser, B. Plattner: An Error Tolerant, Scalable Video Stream Encoding and Compression for Mobile Computing, ACTS Mobile Summit 96, Granada Spain, Vol. II, pp. 762-771, November 1996.

[DWDA97] D. Decasper, M. Waldvogel, Z. Dittia, H. Adiseshu, G. Parulkar, B. Plattner: Crossbow - A Toolk it for Integrated Services over Cell-Switched IPv6; Proceedings of the IEEE ATM'97 Workshop, Lisboa, Portugal, June 1997.

[EdMV95] R. Edell, N. McKeown, P. Varaiya: Billing Users and Pricing for TCP, IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, September 1995, pp. 1162-1162.

[FSVP98] G. Fankhauser, B. Stiller, C. Vögtli, B. Plattner: Reservation-based Charging in an Integrated Services Network, 4th INFORMS Telecommunications Conference, Boca Raton, Florida, U.S.A., March 1998.

[FaSP98] G. Fankhauser, B. Stiller, B. Plattner, Arrow: A Flexible Architecture for an Accounting and Charging Infrastructure in the Next Generation Internet, Accepted for publication in “Netnomics”, 1998.

[Herz96] S. Herzog: Accounting and Access Control for Multicast Distributions: Models and Mechanisms, PhD dissertation, University of Southern California, U.S.A., 1996.

[HeSE97] S. Herzog, S. Shenker, D. Estrin: Sharing the "Cost" of Multicast Trees: An Axiomatic Analysis; IEEE/ACM Transactions on Networking, Vol. 5 No. 6, December 1997, pp. 847-860.

[HoSp96] D. Hoffman, M. Speer: Hierarchical video distribution over internet-style networks. In ICIP’96, Lausanne, Switzerland, September 1996.

[LiPP97] X. Li, S. Paul, P. Pancha, M. Ammar: Layered Video Multicast with Retransmission (LVMR): Evaluation of Error Recovery Schemes, In G. Parulkar (Ed.), NOSSDAV 97, Springer, Berlin, Germany, Lecture Notes in Computer Science, May 1997.

[MacK97] J. MacKie-Mason: A Smart Market for Resource Reservation in a Multiple Quality of Service Information Network, Report, University of Michigan, September 1997.

[MaVa94] J. MacKie-Mason, H. Varian: Pricing the Internet; In Public Access to the Internet (Kahin B. and Keller J., Eds.), Prentice Hall 1994.

[MaVa95] J. MacKie-Mason, H. Varian: Pricing Congestibel Network Resources, IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, September 1995, pp. 1141-1149.

[McVJ97] S. McCanne, M Vetterli, V. Jacobson: Low-complexity Video Coding for Receiver-driven Layered Multicast, IEEE Journal on Selected Areas in Communications, Vol. 16, No. 6, pp. 983-1001, August 1997.

[RaRK96] S. Ramesh, C. Rosenberg, A. Kumar: Revenue Maximization in ATM Networks Using the CLP Capability and Buffer Priority Management; IEEE/ACM Transactions on Networking, Vol. 4, No. 6, pp. 941-950, December 1996.

[ShCE96] S. Shenker, D. Clark, D. Estrin, S. Herzog: Pricing in Computer Networks: Reshaping the Research Agenda; Communications Policy, Vol. 20, No. 1, 1996; also available at URL
ftp://parcftp.xerox.com/pub;net-research/picn.ps

[Shen95] S. Shenker: Some Fundamental Design Decisions for the Future Internet; IEEE Journal on Selected Areas in Communications, Vol. 13, No. 7, 1995, pp. 1176 - 1188.

[WaKS97] D. Walker, F. Kelly, J. Solomon: Tariffing in the New IP/ATM Environment, Telecommunications Policy, Vol. 21, pp. 283-295, May 1997.