A Charging Model for Sessions on the Internet

Nadia Kausar, Bob Briscoe[1], Jon Crowcroft
Department of Computer Science, University College London
Gower street, London WC1E 6BT, UK
[1]BT Labs, Martlesham Heath Ipswich IP5 3RE, England
n.kausar@cs.ucl.ac.uk, rbriscoe@bt.com, jon@cs.ucl.ac.uk

Abstract

A chargeable session on the Internet may consist of more than one underlying chargeable service. Typically there will be two, one at the network layer and one at the session layer. Since different applications can have different demands from the Network, a generic charging scheme has to separate the service provided by the network from the service provided by an application/service provider. In this paper we propose a pricing model which is session based and we look at the impact of this on real-time multimedia conferencing over the Internet. In this model, we are trying to allow for the optional integration of charging at the network layer with charging at the session layer, while keeping the underlying technologies still cleanly apart. This paper also highlights the fact that the main problem of pricing application on the Internet is not just a simple case of analyzing the most technically feasible pricing mechanism but also making the solution acceptable to users. We take the position that session based pricing is easier for end users to accept and understand and show why this is the case in this paper.

1. Introduction

As the popularity of the Internet grows, the number of services offered over the Internet grows with it. Users normally pay a flat fee to join the Internet and are forced to get used to a service which is not guaranteed and prone to variable delays. However, different applications have very different service requirements. For instance, some applications like email, can tolerate significant delay without users expecting discernible performance degradation, while other applications, such as audio and packetised video degrade perceptibly with even extremely small delays[4]. With rapidly diverging types of tasks the need for traffic characterization on the Internet is becoming very obvious. Cocchi et al[4] have argued that, in

order to produce performance incentives it is necessary to support service-class sensitive pricing for any multi-class service discipline. Using this paradigm it is possible for the user to prioritize their applications to conform to what they perceived to be acceptable QoS values. In this situation the user has an option to pay a higher price for the higher quality.

The services on the Internet have a two level matrix for charging. One is from the application perspective, and the other is from the network perspective. Research has been done to provide better than best-effort QoS in the network and provide a corresponding charging model for the added QoS (e.g., charge for throughput, bandwidth, delay etc.). Whereas, Application related pricing, i.e., charging certain fee for an application has been left to different application/service providers. These applications can have either a fixed fee or could be usage based (i.e. charge the users if they have used the application over a certain time period). In this paper, *the term "pricing" is used to refer to the process of setting a price on a service, a product, or on content*. Whereas, *"charging" determines the process of calculating the cost of a resource by using the price for a given record, i.e. it defines a function which translates technical values into monetary units*[17].

As previously mentioned, different application can have very different demands from the network Therefore, in order to provide a comprehensive service for an application, a user must be able to deal with separate charges for both the network and the application QoS. For example, in a video conference, participants may just want to listen to a conference and may not require a guaranteed bandwidth. In this case, these users can be charged to join the conference (e.g. obtain password to join the conference) but pay nothing for reserving network resources. However, if the network resource is scarce then a price will combine both the (minimum amount of)

network resource required to transmit the conference and the facility to join the conference (obtaining passoword – an access key).Therefore, it is best to provide a charging scheme that is not directly integrated with network QoS and specific applications. *In this paper, we propose a session based pricing, we use the term "session" to define the lifetime of activities of a single/group of users.* The aim is to provide protocol independence, in the sense that, different sessions(e.g. multimedia conferencing, multiplayer games or e-commerce activities) from the application layer can be charged independent of any different basis for charging network resources.

A number of approaches have been proposed for control of usage and explicit allocation of resources among users in time of overload, both in the Internet and in other packet networks[2]. RSVP [13], in combination with the Integrated Service model, can be used to explicitly reserve a path or flow between end points in a network Recent research has focused on a more generalized means of providing network QoS based on tagging packets where 'out' tagged packets receive congestion indication first and will be dropped when congestion occurs(diff-serv)[3]. The goal of session based pricing is to allow Internet service provider (ISP) to charge applications that can use a variety of network reservation mechanism; such as RSVP, diff-serv, or DRP[18].

Note that, a session can use multicast to achieve N-to-N communications at the network layer, or it could use an IPtelephony gateways to interwork PSTN phone sets with an IP based conference. Therefore, this paper looks at the issues concerned with commercial model for Multicast, different service aggregation model and outlines the impact of session based pricing on it. The sections in this paper are organised as follows: section two reviews different basis for charging in the network layer, section three looks at various ways services can be aggregated(bundling) for session based pricing, section four looks at the design approach for session based pricing, section five highlights the commercial model for multicast, and the last section concludes the paper.

2. Review of basis for charging in the network

The main reasons for charging on the Internet are: a) to cover the cost for providing the service by service providers b) make a profit (Service providers) c) control the behaviour of users or limit the usage to benefit higher paid traffic. Different mechanisms have different types of technical and economical advantages and disadvantages. In [5], it was shown that users repressed their usage of the network when faced with usage-based charging. The complexities of understanding the criteria the users are paying for have an affect on payment as well. That is to say, if a user is presented with a complex bill that shows different criterias, and how different schemes they subscribed have different prices, there is a likelihood the user will prefer the flat - rate option.

The charging policy in telephone networks has existed for a long time and it works very well. Telephone companies offer a menu of local calling plans, some usage-based (e.g., metered service), some capacity based (e.g. unlimited service), and some a combination of both (e.g. a certain number of free minutes per month, plus a metered rate for calls in excess of this number). It is likely that the same will happen in computer networks, with some users choosing usage based and others choosing capacity based charges, and many being somewhere between[15].

The two most discussed pricing schemes which can be implemented vary easily for the Internet traffic are a) Capacity pricing and b) Usage based pricing. In **capacity based pricing**, a user would purchase a profile, called an expected capacity profile, based on the general nature of his/her usage. For example, a user exploring the web would need a very different profile from a scientist transferring a sequence of large data sets[2].Expected capacity pricing has the advantage of stable budgeting for network use. Also, expected capacity gives the providers a more stable model of capacity planning. If users are permitted to install and use different profiles on demand, the provider must provision somewhat more conservatively, to deal with peaks in demand. However, the biggest drawback of this scheme is that to this point the description of bandwidth allocation has been in terms of sender of the data, when the sender may be generating data because the receiver initiated it (e.g. in ftp case, where the server may be sending data when the user has requested the file).

In **usage based pricing**, the users pay for the volume of traffic (as well as length of time) they are interested in. The argument could be that if the resource is limited and the existing resource are used in different ways, service classes could be applied to differentiate its use appropriately. The biggest argument against this scheme is that usage based charges change the user perception and may decrease user's usage.

Edge Pricing Shenker et al have suggested a method to price the traffic where congestion costs are estimated using the expected congestion (e.g. time of day) along the expected path. Therefore, the resulting prices can be determined and charges are assessed locally at the access point (i.e. the edge of the provider's network where the user's packet enters), rather than computed in a

distributed fashion along the entire path. Edge pricing has the attractive property that all pricing is done locally. Interconnection here involves the network providers purchasing services from each other in the same manner that regular users purchase services[15].

Paris Metro Pricing (PMP) Another way to deal with congestion in packet networks is provided by the PMP model[11]. Odlyzko suggests that an end-user should be required to pay more to use a particular queue, although its architecture would be identical to a cheaper queue.

than the valuation of service itself. The other problem is the bid is on a per-packet basis, yet many applications involve sequence of packets. It is impossible to independently set the valuation of a single packet in a file transfer, when the true valuation is for the set of packets.

3. Entities that bundle services

When a user is choosing a session based price,

Fig. 1. Bundling services
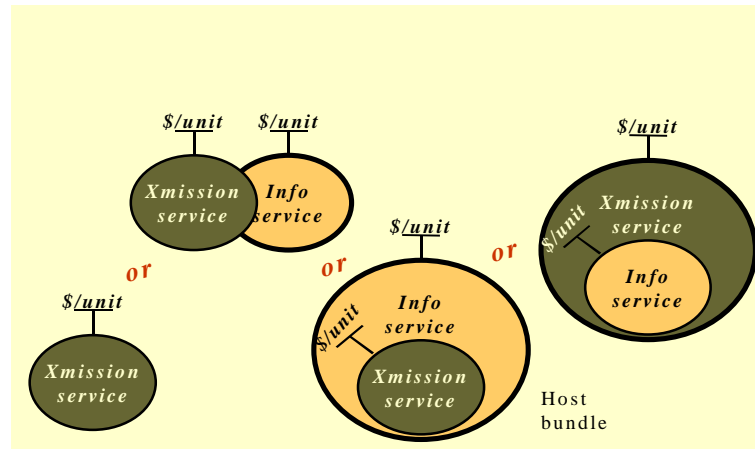
The idea is that the queue that is more highly priced would attract less traffic and therefore suffer from less congestion than the queue with the lower price. PMP does not deal with more than one dimension of QoS. There would need to be a number of bands for each combination of bandwidth differentiation, latency differentiation and reliability differentiation. It is not true that all high bandwidth application also needs high reliability and latency.

Smart Market proposal One of the most ambitious pricing proposals for best effort traffic is the "smart-market" proposal for Mackie-Jason and Varian described in [9].

In this scheme, each packet carries a "bid" in the packet header; packets are given service at each router if their bids exceed some threshold, and each served packet is charged this threshold price regardless of the packet's bid. This threshold price can be thought of as the highest rejected bid; having the packet pay this price is akin to having them pay the congestion cost of denying service to the rejected packet. This proposal has stimulated much discussion and has significantly increased the Internet community's understanding of economic mechanisms in network. However, there are several problems with this proposal[15]. The biggest problem associated with this scheme is that submitting a losing bid will typically lead to some unknown amount of delay (since the packet will be retransmitted at a later time), rather than truly not ever receiving service, so the bid must reflect how many utility loss this delay would produce rather

they can be offered different services with different ways to pay them. Bundling is the service aggregation between different entities. It is a business choice made by the services(providers), made for commercial reasons (e.g. profit opportunity or simply wanting to offer a more useful service). As shown in Figure 1, the transmission service and information service are bundled together (marked as host bundling)where the user pays a certain fee for (example downloading video) the information service provider, who in turn pays the network provider for providing the transmission facility. The user is not aware that information provider is paying a fee for the transmission service.

From a research perspective, there are mainly two types of users: advanced users and novice users[Bouch98]. In the former case, users tend to have theoretical knowledge of networking environments, familiar with syntactic aspects applicable to real-time and data-driven tasks. In the latter case, novice users are mainly the type who have little or no theoretical knowledge of networking environments, unable to directly map technical syntax onto underlying conceptual consideration.

These types of users in the target market make a difference as to how services are bundled.The framework for charging to use different services with different quality on the Internet is quite complex. The parameters charged can be very dynamic and variable. For example, in the telephone system, all calls require the same

network capacity and the same quality of service, whereas flows in the Internet can differ widely in their need for capacity, control of delay, or other features. Especially in the context of associating value with enhanced services, it must be possible for the users to describe the service they require. The features that can be charged are: throughput, speed, accuracy(assertions connecting QoS to the ability of the search engine for example to deliver the requested information), accessibility and reliability. Therefore, for the novice type of users, there may be a requirement to set the session price for them, otherwise they have to be educated through the process of quality of service aspects of the Internet.

In this section we look at different service aggregation methods, i.e. various ways a user can be billed for a particular service he/she used over the Internet. These mechanisms have to be easily understood by the people so that they will be interested in using them. We would like to propose that there will be mainly two ways to bundle.There are:
1. ISP bundling
2. Session owner bundling
And there is always another approach which is based on
3. User's choice

1. ISP bundling – In this scenario, the ISPs will set a price for a given session and the hosts and the participants directly pay their ISP for that session. This is probably not a very attractive option for the ISPs because they have to work out separate prices for interconnecting with different providers for each session, how many people possibly want that service and work out a set price for every session they are providing. ISPs providing Internet telephony services should pay access charges to the local telephone companies as do other long-distance service providers. However, the ISPs can actually make a sufficient amount of profit by providing a price on session based, because a lot of users/hosts do not actually want to go through the trouble of working out a price for a session. In order to bill the users for that session, the ISP has to take into consideration that users may pay into their account (which may/may not exist , so for every session they have to create separate billing account) or by credit card or by e-cash. As mentioned that it may not be the most attractive option for ISPs.

ISPs will have policies which can be exchanged among policy-enabled entities. DIAMETER[14] is currently a proposal which for example, can be used for ISP bundling. It is designed as a common platform for several Internet services, such as

AAA(authentication, authorization and Accounting), network-edge resource management and VPN (virtual private network). So for example, when a caller (e.g. a SIP proxy server) is being notified to set up a call for a user, it first initiates a DIAMETER request command to its policy server with all the information about the user. The server, in turn, checks the request against admission control policy database, and returns the findings in a DIAMETER response message[12]. [12] attempts to cater only for ISP bundling, but it is unlikely DIAMETER would be the preferred solution for non-ISP bundling. Therefore a more general solution would be beneficial.

2. Session owner bundling – In this scenario, the master of ceremonies (e.g. an organiser of a conference) sets the price for individual user or mainly an organization, where the novice users do not have to know the implications. For example, user A decides to host a conference in UCL, 1999 for 2 days. This conference requires access to the Mbone[10] in order to multicast the session. So the host has to work out what is the minimum bandwidth required to transmit video(frame rate) and in order to deliver audio what is the bit rate required. After that the session owner sets a common price for all that absorbs and hides peaks and throughs in costs for each participant. A slight premium allowance above the expected average cost involved underwrites the host's risk. This might either turn a profit for the host or be returned to all participants in equal shares (co-op dividend). Each participant's cost to the host will depend on their ISP's price, but the host is wholesaling (hiding) this to participants. This may be a lot of work for the host to work out a suitable price. This scheme will be attractive for a type of host who holds a lot of sessions like that a year and the host is likely to be a big organization. As for the user is concerned, they do not have to worry about the technical aspects of the conference and it makes it definitely simple for them to just pay the host and participate in the conference. The question remains, will a user be interested in paying a fixed amount for which they are confined to the policy the host/session owner has set?

3.   User - In this scenario, the user has the choice to go either with the "best-effort service" for a session or can pay their ISP directly for guaranteed service. Normally for all of the above option as well, the frame rate for video and for audio the required bandwidth will be advertised on SDR(see section 4 for further discussion). Therefore, it is up to the user to pay a certain fee

for a certain amount of guaranteed service. For novice users, they do not necessarily need to know the technical details. There will be the option in the form of a sliding bar marked with values (either monetary values or other forms of prices), and increasing the value of the sliding bar will increase the quality.

With this option, the host or the ISP do not have to set certain prices for everyone for different sessions. Also, it gives the user the flexibility to go with their own policy, i.e. they are not confined to ISP's or the host's policies. For all of the models of payments above strong security is necessary both between routers and policy servers and between policy servers and the billing system that connects policies to economics because their interaction implies financial transactions. Whatever the bundling scenario is and whether an ISP or a user is setting the price, they can use a session based pricing interface (as discussed in the section below) to serve their purpose.

4. Model for application driven session pricing

An important aspect of the problem of designing a model to price real-time applications on the Internet is that the Internet architecture is based on the network layer not knowing the properties of the applications implemented above it. Therefore, in this model the knowledge of underlying resource management and network implications of providing a guaranteed service is not necessary and has been separated from the applications. ISPs or the bandwidth broker sets a certain price for each session that can be accessed from the session layer.
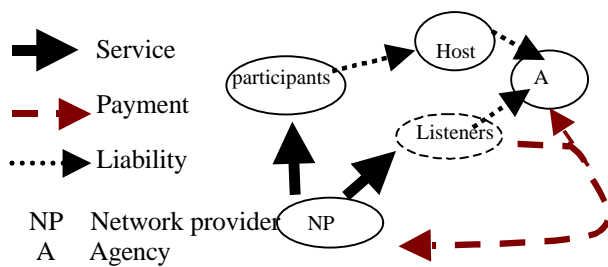

Fig. 2. Model of interaction between participants and agency

The design philosophy of this model is quite simple. Let us take a multimedia conference for example, there will be few participants among which some are just listeners. This session is advertised by some arbitrary means (e.g. SDR [Kirst97] or a web page), with the session's price While in this paper we have focused on monetary values to participate in a session, the underlying

accounting structure and pricing architecture should allow the use of these other incentive forms if they are locally applicable.

As discussed in section 3, the user driven system where the user has the choice of either paying a certain fee for guaranteed QoS or not paying. So there will be an "agent" who will be responsible for collecting the payment. The session based pricing comprises a "back-end" , whose job is to inform the service provider or the initiator (depending on who is charging and what the policy is) that the specific session is being paid for and a guarantee for that service for that price is required. Each router, on receiving a packet, must able to determine whether the router is within the paid region. There are only two ways that a router can have access to information about a flow. Either it is stored in the router (this is not the preferred option), or in the packets of the flow.

The second part is a "front-end" which allows a client to provide inputs in the selection process. In this scenario we have used multimedia conferencing as an example where there are different classes of participants. So the participants who are just listeners can choose to pay a flat fee whereas a speaker will pay an additional amount for that session. However, for example, if the speaker is an invited speaker then he/she may pay nothing. Floor control [7] for the session plays a very useful part for this pricing scheme.

If a user initially chose the option not to speak then the floor control option is not enabled. However, we realized that a listener may have a question at the end of a session, but the amount of traffic that will be generated by this question may have an impact on the network if the resource available is scarce. For most of the existing conference tools there is a facility to use the chat option where the users can type in their question.

SDR[8] is used to advertise multimedia conferences, and to communicate the session addresses (whether multicast or unicast)and conference-tool- specific information necessary for participation. This would be an ideal tool to advertise the prices associated with the sessions. Currently the user interface looks like as shown in Figure 3.
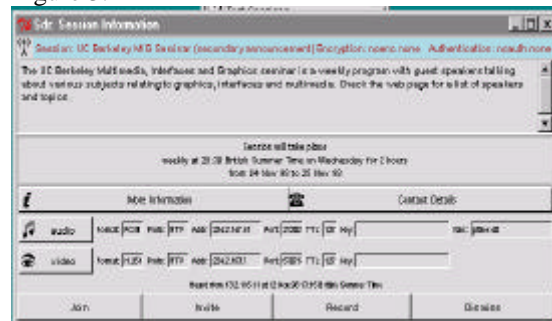

Fig. 3. Session directory would need an option for payment

5. Support from Network layer

Elastic   Inelastic

Email   File transfer   WWW   Streamed   Interactive Apps
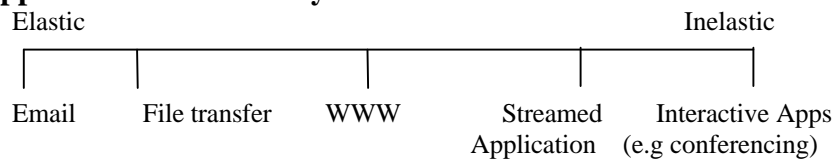   Application   (e.g conferencing)

Fig. 4. Relative traffic elasticity

The Internet today offers a single class of service, where all packets are serviced on a best effort, First-in–First-Out (FIFO) basis. Disrupted audio and video due to packet losses make multimedia conferencing less efficient and less effective. The applications that generate traffic can be located on a continuum (see Figure 4) which also represents the delay tolerance. As the amount of real-time traffic increases there may be a corresponding need to define a richer set of QoS parameters for these traffic types[1].

Although users are normally prepared to put up with delay with elastic applications because it is expected to be delivered later in the day and picked up some other time, one may send an urgent email which can be treated as a real-time or inelastic application (for example, an email informing someone to join a conference immediately).

In order to guarantee the service that is chosen from the session based application pricing interface, the network has to provide enough resources. Since an RSVP API[16] is currently available, we suggest to integrate RSVP with the different models of session based pricing. However, as discussed in section 2 there are other ways to reserve the resources or characterise the packets that are being paid for in network layer. In this paper, we are not focussing on any particular charging scheme for network or service, any number of combinations can be used. We are assuming the commercial application will be paid for and the underlying resources will be reserved or characterized in a way that will support the application.

To ensure voice and data are being delivered properly, users can make the use of end-to-end resource reservation protocols to set up reserved "flows". Another alternative is to mark the packet header as "premium service" so that they can be delivered with low delay and rate guarantees inside the network. Both approaches imply that the network-edge routers may need to interface with policy servers to manage link resources.

6. Multicast Model

It is held that multicast offers significant advantages to the Internet community. Multimedia real-time applications which are being multicast pose more of a challenge to be priced and different access rates need to be considered carefully when pricing the senders/receivers. A multicast address is merely a logical name, and by itself conveys no geographic or provider information. Multicast routing identifies the next hop along the path for packets arriving at an interface, multicast routing does not identify the rest of the tree. Thus, estimating costs in the multicast case requires an additional piece of accounting infrastructure. One approach to price the receivers is to introduce a new form of control message – an accounting message – that would be initiated when the receiver sends its multicast join message[15]. These accounting messages would be forwarded along the reverse trees towards each source, recording the "cost" of each link it traversed and summing costs when branches merged.

With the User bundling scenario, the session based pricing solves the problem of charging receiver/sender in a multicast session. As discussed previously, the user can pay a set price regardless their position in a multicast tree. If the receiver wishes to receive a session with a certain guarantee, they just have to pay. In user bundling system, the price to be paid for a session's quality is upto the user, so whether the user is a multicast receiver or not, does not really affect the pricing scheme. If the multicast tree is organised in a hierarchical structure, then the host or the ISP (if it is a host or ISP bundling system being used to pay for services) can negotiate or set a price for a particular branch of the tree. Then, if one of the child nodes join a session which needed to be paid for, the node can obtain the price from the nearest parent node.

Another issue to be addressed is: which party (content provider, ISP or receiver) does multicast transport offer the most intrinsic value compared with unicast transport? In overall, one can say that multicast access and peering agreements are likely to be placed on a very different financial basis from the existing unicast agreements. the multicast sender (e.g. content provider) benefits greatly from multicast, since

access costs are drastically reduced. There is little multicast benefit to the receiver. To the receiver it makes little difference whether multicast or unicast is used (assuming, received bandwidth is charged at the same rate unicast or multicast). By default, the ISP should charge multicast senders (e.g. content providers) more for multicast access bandwidth (sent into the network) than for unicast access bandwidth.

7. Conclusion

Different types of traffic sent into the network may have different QoS requirement associated with them. The satisfaction a network user derives from their network access depends on the nature of the application being used and the quality of service received from the network. Since the nature of the Internet architecture is based on the network layer not knowing the properties of the applications implemented above it, we have proposed a session based pricing model that operates over existing network reservation/pricing schemas and

Setting a session price that will profit the ISP or the content provider, and yet still be price-competitive with their competition, can be difficult to predict. The complexity to predict and implement a profitable price for session based pricing is still an open issue and a subject for further research. However, session based pricing has the attractive features of providing a more direct way of communication costs to the user and the flexibility to implement it with any different basis for charging network resources. The proposed work is currently under implementation.

References

[1] Bouch A., " A user cnetered approach to the design and implementation of Quality of service and charging mechanisms in Wide-area Networks" – 1st year report, http://www.cs.ucl.ac.uk/staff/A.Bouch

[2] Clark D. "Adding service discrimination to the Internet " September 1995, presented at MIT workshop on Internet Economics

[3]Clark D.(MIT), A model for cost allocation and pricing in the Internet, presented at MIT workshop on Internet Economics, Mar 1995 "http://www.press.umich.edu/jep/works/ClarkMod el.html"

[4] Cocchi R., Shenker S., Estrin D., Zhang L. "Pricing in computer Networks" – Motivation, formulation and Example , IEEE/ACM Transactions on Networking, vol. 1, Dec. 1993.

[5] Cosgove J., Linhart P. "customer choices under local measured telephone service" Public utilities fortnightly, 30, pp 27-31, 1979

[7] Kausar N., Crowcroft J. – " Floor control requirements from reliable IP multicast" 8th IFIP Conference on High Performance Networking (HPN'98) The Millennium Push of Internet, Vienna September 21-25, 1998

[8] Kirstein P., Whelan E. "SAP - Security using public key algorithms"Internet draft http://www.ietf.org/internet-drafts/draft-ietf-mmusic-sap-sec-04.txt

[9] Mackie-Mason J., Varian H. "pricing the Internet" – In brian kahin and James Keller, editors, Public access to the Internet. Prentic – Hall, New Jersey 1995, URL: ftp://gopher.econ.lsa.umich.edu/pub/papers/Pricin g_the_Internet.ps.Z

[10] Mbone information site, http://www.mbone.com/techinfo/

[11] Odlyzko A. "A modest proposdal for preventing Internet congestion" 1997, http://www.research.att.com/~amo/doc/recent.html

[12]Pan P., Schulzrinne H. "DIAMETER: policy and Accounting Extension for SIP" Internet draft, Internet Engineering task Force, July 1998

[13] Internet draft –A Framework for Use of RSVP with Diff-serv Networks http://search.ietf.org/internet-drafts/draft-ietf-diffserv-rsvp-01.txt

[14] Rubens A., Calhoun P "DIAMETER base protocol" Internet draft, Internet Engineering task Force July 1998

[15] Shenker., Clark d., Estrin D., Herzog S. " Pricing in computer networks" ACM Computer Communication Review, vol. 26, pp. 19-43, Apr. 1996.

[16]Stevens "Advanced Programming in the Unix environment"

[17] Stiller B., Fankhauser G. "Charging and Accounting for Integrated Internet Services – state of the art, problems and trends", INET 1998, Switzerland, July 21 – 24, 1998

[18] White P. and Crowcroft J.. A Dynamic Sender-Initiated Reservation Protocol for the Internet. 8th IFIP Conference on High Performance Networking (HPN'98) Vienna September 21-25, 1998