# Modelling and analysis of Internet Pricing: introduction and challenges

Yezekael Hayel[1], Patrick Maillé[2], and Bruno Tuffin[1]

[1] IRISA/INRIA
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France
(e-mail: yhayel@irisa.fr, btuffin@irisa.fr)

[2] GET/ENST Bretagne
2 rue de la Châtaigneraie
35576 Cesson Sévigné Cedex, France
(e-mail: patrick.maille@enst-bretagne.fr)

**Abstract.** This paper/presentation aims at introducing the reasons why switching from the current flat-rate Internet pricing to another scheme is required, at briefly classifying the existing propositions, and at highlighting the challenges that still have to be tackled in the area. Pricing has indeed become a hot topic in the networking literature in order to control congestion, differentiate services among users and somehow fairly share the resource, but is still the subject of debate about how, and even if, it should be implemented.
**Keywords:** Pricing, Game theory, Modelling, Optimisation.

## 1 Introduction: why changing?

The Internet has experienced a tremendous success during the last decade. Starting from an academic (and somewhat free) communication network, it has been expanded to commercial purposes. The way customers are currently charged is based on a so-called *flat-rate* price: they pay a fixed subscription fee to an Internet Service Provider (ISP) and have an unlimited access to the network.

Due to the success of this expansion, the amount of Internet traffic has soared in an exponential way, from the increase in the number of subscribers, but also from the more and more demanding applications used by customers, in terms of bandwidth, but also in terms of quality of service (QoS) requirements. Indeed, the proportion of telephony, video and multimedia traffic for instance is increasing with respect to data file transfer and email.

This traffic growth and diversity has highlighted the following problems of the flat-rate pricing scheme, which may therefore have become irrelevant:

1. congestion is observed, resulting in erratic QoS: longer delay, larger jitter and increase of losses. Some people argue that increasing capacity can solve the problem and that, thanks to optical fiber especially, we are safe for a while [Anania and Solomon, 1997]. This is actually the topic of the

lasting debate around pricing in the networking community. We indeed believe that this argument may be true for the backbone network, but it seems unlikely to switch from copper lines to optical fiber in access networks, due to a high cost, issue also known as the *last mile* problem [Bernstein, 1997]. Moreover, in wireless networks, capacity (the radio spectrum) is and will probably remain limited.

2. Next, a flat-rate pricing is an incentive to overuse the network: any selfish user has interest in consuming as much as possible, whatever the loss of QoS imposed on other users is. The charge is thus unfair since small users pay as much as big ones.
3. Finally, a flat-rate pricing does not allow service differentation among users (and applications), since everybody is served at the same level, with therefore the same QoS.

As a consequence, designing a new pricing scheme is probably the most simple and natural way to cope with congestion, control demand, fairly share resources and differentiate services among users and/or applications [Courcoubetis and Weber, 2003]. The following issues have to be addressed in the design process: which families of new pricing schemes could be used (Section 2)? What are the externalities imposed on other users that have to be dealt with (Section 3)? What modelling tools and properties need to be verified (Section 4)? How do users react to prices (Section 5)? What is the trade-off between mathematical efficiency and engineering feasability (Section 6)? Section 7 also briefly addresses a new challenge in the pricing community: how do independent ISP will exchange traffic and how will they charge each other?

## 2 Changing to what?

Changing the simple flat-rate pricing scheme to a usage-based or congestion-based one appears thus preferable to us. Some may be worried about the acceptance of such a move, due to the current strong public preference for flat-rate, but it is likely that people will eventually get accustomed to it. Note that sophisticated pricing schemes already exist in other areas such as airfare rate [Odlyzko, 2000] or the new London city toll pricing for instance.

There is a broad range of new schemes proposed in the literature. We can sort them into:

1. pricing schemes for guaranteed services through resource reservation (using RSVP protocol for instance) and admission control (the reader may see [Paschalidis and Tsitsiklis, 2000] or [Songhurst (ed.), 1999] for instance[1]).

[1] Note that the references throughout the paper are not exhaustive but try to be as representative as possible.

2. A promising proposal, called Paris Metro Pricing [Odlyzko, 1999], consists in partitioning the network into several logical subnetworks, each subnetwork working as the current one, but with different access charges, so that the most expensive ones would likely be less congested. Unfortunately, this proposal has been shown to be inefficient in a competitive context [Gibbens *et al.*, 2000].
3. Another quite simple scheme is the so-called Cumulus pricing analysed in [Reichl and Stiller, 2001,Hayel and Tuffin, 2005a] where positive or negative points are awarded depending on the respect of the predefined contract, and contract renegociation (with penalties) is periodically applied.
4. Priority pricing [Cocchi *et al.*, 1991] among different classes (at the packet level) is probably the scheme which fits the most directly the proposed DiffServ architecture. This scheduling policy has nevertheless been compared with other policies such as generalized or discriminatory processor sharing [Hayel *et al.*, 2004] [Hayel and Tuffin, 2005b] when corresponding optimal prices are used. Also, priority for rejection at buffers implementing active queue management has been studied in [Altman *et al.*, 2004].
5. Auctioning, either for priority [Marbach, 2001] or for a proportion of bandwidth [Semret, 1999] [Maillé and Tuffin, 2004] has also recently received a lot of attention.
6. Finally, a last main group is dealing with pricing based on transfer rates and shadow prices, following. the tremendous work of Kelly *et al* [Kelly *et al.*, 1998].

## 3 Technologies and externalities: what to price for?

In communication networks, selfish behaviours lead to unsatisfactory outcomes because of *externalities*: the value a user gets from the network depends on the other users. As an example, in a problem of bandwidth sharing on a communication link, a user that is allocated an amount of bandwidth prevents the others from obtaining that resource, and some requests may be rejected. The externality can thus be defined as the loss of valuation a user's presence imposes on the others.

In order to drive users to behave in a more efficient way, externalities have to be taken into account when designing a pricing mechanism. Notice that externalities are often negative, but can also be positive in some cases: the most classical example in resource allocation problems is the case of multicast communications, where several users interested in the same flow have a common interest and therefore an incentive to cooperate. However they still compete against users interested in other flows.

Externalities may take different forms depending on the *technologies* used and the *performance criteria* users are sensitive to: a user willing to transfer a file will be sensitive to the entire transfer duration (losses inducing retransmissions), whereas for some real-time applications delay is the most important constraint, few losses being permitted. Considering wireless ad hoc networks transmission rates and battery consumption [Crowcroft *et al.*, 2003] are additionally critical.

To analyse properly the externalities, the mechanism designer has first to identify the limited resource, that can be bandwidth or computing capacity for wired networks; spectrum, battery and/or transmission power for wireless networks. Then the correlations between the relevant performance criteria can be studied as a function of the limited resource usage. To that extent, the technological specificities of the systems and protocols should be considered. In wireless networks for instance, the way multiple access is provided (Code, Time and/or Frequency Division Multiple Access) has an influence on the externality impact, since it determines the form of the interference that affects the performance criteria through the signal to noise ratio. For real-time applications, the scheduling policies implemented in the different nodes of the network are critical, since they highly influence the overall transmission delay.

For each performance criterion the designer focuses on, and for each communication system, externalities may have a different form, and modelling and studying them raises different problems. One important stage of mechanism design is to carefully study those problems, in order to build the right incentives (through prices) to drive the user behaviour to the desired direction.

## 4 Mathematical tools and properties involved

A pricing mechanism can be justified by its properties in terms of some economical criteria, such as *efficiency* (maximization of social welfare), *fairness* [Kelly *et al.*, 1998], maximization of network revenue or of the number of accepted clients... Such results need the outcome of the game to be foreseen, which implies that the user behaviour has to be predicted.

Actually, the study of users reactions to a pricing mechanism usually relies on selfishness: the users are expected to act so as to obtain the highest utility, regardless of the consequences on the others. The theoretical framework to study such problems is *game theory* [Fudenberg and Tirole, 1991], and more precisely noncooperative game theory[2]. When the mechanism is well designed, there exists a unique Nash equilibrium that predicts the outcome of the pricing game.

Game theory often implies the use of optimization. Indeed, optimization occurs at different levels:

[2] Game theory also includes the study of cooperative games, however in the context of communication networks it is not likely that users know each other and have an interest in cooperating.

- users try to maximize their utility at the outcome of the game. Depending on the problem considered, that optimization may rely on queueing theory (when delays and losses at the network nodes are the externalities), signal processing (especially in wireless networks) or other mathematical modelling tools adapted to the considered network. An important and interesting property in many pricing schemes, called *incentive compatibility*, states that a user cannot do better than following the designer point of view, that is revealing his real willingness-to-pay for quality of service or choosing the proper class in multiclass systems for instance.
- At the mechanism designer level, since the optimization from the user point of view can be predicted (from what is said in the previous item), the Nash equilibrium can be oriented to a point optimizing some desired criteria.

## 5 User behavior modelling

As introduced in the previous section, modelling the users' valuation of service is required and is one of the main issues of Internet pricing. Users' preferences (or levels of satisfaction) are expressed by functions called *utility* functions [Fudenberg and Tirole, 1991]. In most Internet pricing papers, the inputs of these functions are the throughput or used bandwidth, the average delay or loss ratio, more generaly the considered externality, and may depend on the type of application. In the literature, the utility functions are selected to model the real user behaviour as closely as possible, but also to verify interesting mathematical properties. Those properties are usually the continuity, differentiability and concavity, to make sure that optimal points exist and are unique [Kunniyur and Srikant, 2003].

Nevertheless, one main challenge is to determine a realistic expression of the utility function (or its distribution over a population). For real-time applications for instance, one would expect non-continuous functions, with thresholds under which the utility of being served becomes null. Very few attempts have been published to solve this question. The only cases we are aware of are as follows. In [Beckert, 2000], the utility function is modelled by a Cobb-Douglas function, which requires the determination of several parameters. These parameters have been estimated using a large-scale experiment testing user behaviour which has been performed at UC Berkeley, called the INDEX project [Edell and Varaiya, 1999]. Another worth-mentioning paper is [Gupta *et al.*, 1998], where a quasi-bayesian update algorithm is developped, aiming at estimating the users' waiting cost per unit of time. This approach can be used to estimate the demand elasticity with respect to prices.

To sum up the section, the choice of utility function has a major impact on the pricing scheme analysis, and should be based not only on mathematical interest, but on practical reality (a usual trade-off in modelling). We now deal with another important trade-off between mathematical and engineering efficiencies.

## 6 Trade-off between mathematical and engineering efficiencies

Indeed, to obtain a more efficient model, it is often required that prices react dynamically and instantaneously to an externality evolution, so that the system can be continuously kept at its optimal point. Nevertheless, this requires an important signalling overhead, and is difficult to implement from an engineering point of view (at this point, it is important to emphasize that a main reason of the Internet success is its simplicity, which has to be preserved). It is also important to note that, following the previous section, even users are skeptical with respect to a dynamic pricing, as highlighted by the INDEX project [Edell and Varaiya, 1999]. Again, those trade-offs are important issues a designer has to cope with.

It is interesting to note that a good approximation to dynamic pricing is *time-of-day* pricing. It has been shown in [Paschalidis and Tsitsiklis, 2000] that it leads to an asymptotically efficient scheme, while being simple from an engineering and user point of view. Time-of-day pricing is popular in many areas such as telephony, airfare (where it is rather time-of-year)...

The efficiency problem can also be placed at other levels:

- from a mathematical point of view, the efficiency is more easy to reach if charges are applied at each node of the network (or at least for the whole path). This again induces a signalling overhead in terms of accounting (for total charges have to be computed before being billed to the users), but also requires to inform the user in order to make him accept the transaction. A simpler trend is to charge users at the edges of the network, even if it seems difficult to abstract the network status at the edges in an efficient way (especially if the considered traffic does not pass through the existing bottlenecks).
- Also, applying resource reservation (that is making sure that when your session is accepted, you will get a given QoS for the whole duration of your connection) is appealling mathematically and from the user side point of view, but is intricate to apply to a large network of the Internet size. Scalability is thus the reason why the IntServ architecture initially proposed for Internet QoS has gradually taken place to the DiffServ proposal, where no strict reservation is applied.

## 7 A new challenge: inter-providers pricing

A pricing game that even the opponents of Internet pricing admit to be mandatory is pricing among ISPs in order to deliver their own traffic. Indeed, concurrent ISPs are in competition in the Internet and have to meet traffic forwarding agreements in order to convey their messages to destination if it is not in their network.

A natural way to apply this is to implement auctions between providers and to use, and extend, the Border Gateway Protocol (BGP) usually applied for routing [Feigenbaum *et al.*, 2002]. The goal is then to find lowest-cost routing for sending traffic from an ISP to another that is not directly attainable thanks to BGP. By using VCG auctions, incentive compatibillity can be obtained.

Similarly, pricing for transiting traffic between ISPs and pricing for customers has been jointly studied in [Shakkottai and Srikant, 2005]. Repeated games are used, and, with threat strategies, optimality is shown in the sense that deviating from the equilibrium makes you suffer the worst possible penalty.

## References

[Altman *et al.*, 2004]E. Altman, D. Barman, R. El Azouzi, D. Ros, and B. Tuffin. Pricing Differentiated Services: A Game-Theoretic Approach. In *Proceedings of IFIP/TC6 Networking Conference*, Athens, Greece, May 2004.

[Anania and Solomon, 1997]L. Anania and R.J. Solomon. Flat- The Minimalist Price. In Lee W. McKnight and Joseph P. Bailey, editors, *Internet Economics*, pages 91–118. MIT Press, 1997.

[Beckert, 2000]W. Beckert. *Stochastic Demand Analysis.* PhD thesis, UC Berkeley, 2000.

[Bernstein, 1997]L. Bernstein. Managing the last mile. *IEEE Communications Magazine*, 35(10):72–76, Oct 1997.

[Cocchi *et al.*, 1991]R. Cocchi, D. Estrin, S. Shenker, and L. Zhang. A Study of Priority Pricing in Multiple Service Class Networks. In *Proceedings of SIGCOMM'91*, pages 123–130, 1991.

[Courcoubetis and Weber, 2003]C. Courcoubetis and R. Weber. *Pricing Communication Networks—Economics, Technology and Modelling.* Wiley, 2003.

[Crowcroft *et al.*, 2003]J. Crowcroft, R. Gibbens, F. Kelly, and S. Östring. Modelling incentives for collaboration in mobile ad hoc networks. In *Proceedings of WiOpt'03*, Mar 2003.

[Edell and Varaiya, 1999]R. Edell and P. Varaiya. Providing Internet Access: What We Learn From INDEX. In *Proceedings of INFOCOM*, 1999.

[Feigenbaum *et al.*, 2002]J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP-based mechanism for lowest-cost routing. In *Proceedings of the twenty-first annual symposium on Principles of distributed computing (PODC'02)*, pages 173–182. ACM Press, 2002.

[Fudenberg and Tirole, 1991]D. Fudenberg and J. Tirole. *Game Theory.* MIT Press, Cambridge, Massachusetts, 1991.

[Gibbens *et al.*, 2000]R. Gibbens, R. Mason, and R. Steinberg. Internet service classes under competition. *IEEE Journal on Selected Areas in Communications*, 18(12):2490–2498, 2000.

[Gupta *et al.*, 1998]A. Gupta, B. Jukic, M. Li, D. Stahl, and A. Whinston. Estimating Internet User's Demand Characteristics. Technical report, National Science Foundation, 1998.

[Hayel and Tuffin, 2005a]Y. Hayel and B. Tuffin. A Mathematical Analysis of the Cumulus Pricing Scheme. *Computer Networks (To appear)*, 2005.

[Hayel and Tuffin, 2005b]Y. Hayel and B. Tuffin. Pricing for Heterogeneous Services at a Discriminatory Processor Sharing Queue. In *Proceedings of IFIP/TC6 Networking Conference*, Waterloo, Canada, May 2005.

[Hayel *et al.*, 2004]Y. Hayel, D. Ros, and B. Tuffin. Less-than-Best-Effort Services: Pricing and Scheduling. In *proceedings of IEEE INFOCOM*, 2004.

[Kelly *et al.*, 1998]F.P. Kelly, A.K. Mauloo, and D.K.H. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49:237–252, 1998.

[Kunniyur and Srikant, 2003]S. Kunniyur and R. Srikant. End-to-End Congestion Control Schemes: Utility Functions, Random Losses and ECN Marks. *IEEE Transactions on Networking*, 11(5), 2003.

[Maillé and Tuffin, 2004]P. Maillé and B. Tuffin. Multi-bid auctions for bandwidth allocation in communication networks. In *IEEE INFOCOM 2004*, Hong-Kong, China, March 2004.

[Marbach, 2001]P. Marbach. Pricing Differentiated Services Networks: Bursty Traffic. In *Proceedings of IEEE INFOCOM 2001*, 2001.

[Odlyzko, 1999]A. Odlyzko. Paris Metro Pricing for the Internet. In *ACM Conference on Electronic Commerce (EC'99)*, pages 140–147, 1999.

[Odlyzko, 2000]A. Odlyzko. The history of communications and its implications for the Internet. Technical report, AT&T Labs, 2000.

[Paschalidis and Tsitsiklis, 2000]I.Ch. Paschalidis and J.N. Tsitsiklis. Congestion-Dependent Pricing of Network Services. *IEEE/ACM Transactions on Networking*, 8(2):171–184, 2000.

[Reichl and Stiller, 2001]P. Reichl and B. Stiller. Edge pricing in space and time: Theoretical and practical aspects of the cumulus pricing scheme. In *Proceedings of the 17th International Teletraffic Congress*, 2001.

[Semret, 1999]N. Semret. *Market Mechanisms for Network Resource Sharing*. PhD thesis, Columbia University, 1999.

[Shakkottai and Srikant, 2005]S. Shakkottai and R. Srikant. Economics of network pricing with multiple ISPs. In *Proceedings of IEEE INFOCOM 2005*, Miami, FL, USA, 2005.

[Songhurst (ed.), 1999]D. Songhurst (ed.). *Charging Communication Networks: from Theory to practice*. Elsevier, Amsterdam, 1999.