

Evaluating Internet Pricing Schemes: A Three-Dimensional Visual Model

Thuy T.T. Nguyen and Grenville J. Armitage

Traditional Internet pricing schemes are coming under continual pressure to adapt to, and encourage, a changing mix of Internet applications and consumer usage patterns. Much research effort over the last decade has been focused on developing more efficient and attractive charging schemes. However, none of the proposed models has been widely deployed. This raises questions regarding the inhibiting factors and missing pieces that make pricing the Internet such a challenge. In this paper, we discuss the problems with current Internet pricing schemes, review the history of Internet pricing research over the last ten years, and summarize the key features and motivations of the most significant models. We develop a novel visual approach to comparing and evaluating such schemes using a three-dimensional (3D) metric encompassing technical efficiency, economic efficiency, and social impact. We address and discuss the important factors that have inhibited the deployment of the reviewed models and suggest productive areas of focus for future Internet pricing research.

Keywords: Internet pricing, evaluation model, 3D.

I. Introduction

Finding a fairer and more efficient charging scheme for the Internet has attracted a significant research effort over the last decade. Many pricing models have been proposed, aiming at an ideal pricing scheme that is able to provide different levels of services to different users with different needs, charge users only for their perceived quality of service (QoS) and consumed resources, support the non-uniformity of Internet traffic with different QoS requirements, and help Internet service providers (ISPs) develop sustainable and profitable business models.

Most proposed models generally achieve a subset of these goals. However, no particular solution has been found to be widely deployed—the Internet is still dominated by flat-rate pricing and simple usage charges (such as charging per volume of traffic or connection duration). This leads to an interesting question: Is the current Internet charging scheme already sufficient? And if not, then what are the missing pieces and inhibiting factors that make Internet pricing still an open question?

In this paper, we place current Internet charging schemes into a broader context of the alternative Internet pricing schemes proposed over the last ten years, summarizing the key features and motivations of the most significant alternative models. We develop a novel visual approach to comparing and evaluating such schemes using a three-dimensional (3D) metric, which encompasses the dimensions of technical efficiency, economic efficiency, and social impact. Finally, we address and discuss the important factors that have inhibited the deployment of those proposed models and suggest some productive directions for short-term and long-term Internet pricing solutions.

Section II begins with a review of current Internet pricing and a discussion of its limitations. An overview of important Internet pricing models investigated over the last ten years is covered in

Manuscript received May 25, 2004; revised Aug. 11, 2004.

Thuy T.T. Nguyen (phone: +61 3 9214 8089, email: tnguyen@swin.edu.au) and Grenville J. Armitage (email: garmitage@swin.edu.au) are with the Centre for Advanced Internet Architectures, School of Biophysical Sciences and Electrical Engineering, Swinburne University of Technology, Melbourne, Australia.

section III. Section IV introduces our visual 3D evaluation model and provides an analysis of the reviewed pricing models using our 3D evaluation approach. Section V discusses the important factors inhibiting the deployment of those proposed models from the points-of-view of both consumers and service providers. Section VI suggests some short-term and long-term Internet pricing solutions before the final conclusion in section VII.

II. Current Internet Pricing Schemes

First we consider the context and limitations of today's Internet pricing schemes.

1. Background of Current Internet Pricing

Figure 1 illustrates the relationship between the major components of today's Internet: end users, ISPs and backbone providers. Most Internet users currently face very similar pricing structures for their usage. Dial-up access to the Internet is often charged at a fixed monthly fee, including either unlimited use or a limited duration connection. A fixed bandwidth connection is normally charged an annual or monthly fee, which allows unlimited usage up to the physical maximum flow rate, dependant on bandwidth capacity. Simple usage-based pricing is also common, in which the users are charged according to the duration of the network connection or the volume of traffic received and sent.

Charging between ISPs is complicated by the lack of a comprehensive inter-ISP settlement structure. Charging arrangements among directly connected ISPs are purely ad hoc. ISPs negotiate bilateral agreements with one another as they see fit. Pairs of ISPs often have peering agreements by which they exchange Internet traffic for free.

ISPs pay to connect to the backbone carriers. Such payments are a simple, low monthly rate depending on the maximum transfer rate of the connection and are usually not based on actual traffic volumes [1], [2].

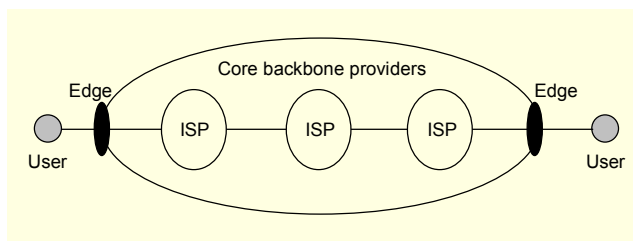


Fig. 1. Major Internet players.

2. Limitations of Current Internet Pricing

Despite the recent increases in available backbone

bandwidth, a major problem that remains is network congestion exacerbated through a user's self-interested behavior. Like all systems with finite resources, the use of a network by one participant can negatively affect the characteristics of the network for others. Such effects might include transient congestion and even packet loss, which is quite noticeable for real-time interactive traffic. In its existing *best-effort* form, the Internet creates no incentive for users to take this problem into account when deciding their usage [3].

Pricing based on connection duration creates no incentive for users to consider the network resources that they consume. Connection duration is also non-linearly related to the consumption of resources in a packet-switched network (for example, browsing for hours at a remote web page that has already been downloaded into an ISP's local cache does not entail much additional network traffic beyond the cache) [3].

Furthermore, new applications with different traffic and service requirements (for example, real-time multimedia applications) are likely to be deployed widely in the future. It is highly desirable for the evolving Internet to meet these increasingly varied service requirements. Flat-rate pricing does not support the diverse nature of emerging Internet traffic sources and might inhibit such development [4].

In addition, there is no clear indication of whether the senders or receivers of data should pay. The proliferation of unsolicited commercial e-mail (often colloquially referred to as "Spam") is a serious problem on today's Internet. A pricing scheme wherein the senders pay the full cost is initially appealing as an effective and simple solution for the unsolicited commercial e-mail problem. However, the simplistic application of such an approach would have a chilling effect on providers of free (or heavily discounted) content over the Web. Clearly, some middle ground is required.

III. Research on Internet Pricing—Scanning the Last Decade

The first attempts to study the problem of charging and pricing for the Internet were undertaken by Cocchi et al. [5] during the early 1990's. In 1994, MacKie-Mason and Varian introduced the idea of using auction mechanisms in "smart markets" for a *best-effort* network. This work was the starting point of applying externalities and congestion pricing to the Internet. In 1995, Shenker [6] formulated an important model based on the integrated services approach that included both soft and hard guaranteed services in the network model. These works provided major stimuli to the research for the following years [7], [8].

The proposed Internet pricing models over the last decade can be classified into two broad categories:

- Pricing for *best-effort* service including *congestion pricing*, *priority pricing*, *Paris Metro*, *zone-based pricing*, and *edge pricing*.
- Pricing with QoS guaranteed, including pricing for integrated services and differentiated services.

The following subsections contain an overview of the important Internet pricing models that have been investigated over the last ten years and have turned out to be of special importance from a practical and economic point of view.

1. Pricing for Best-Effort Service

Among pricing models proposed for *best-effort* service, *congestion pricing* has drawn considerable attention and research effort. From an economic point of view, the use of a network by one user exhibits negative externalities for others—his/her traffic imposing on the network might cause congestion-induced delays or even packet loss to other users' traffic. In principle, such a user should pay the social costs of delaying other users' traffic when the network is congested. On the other hand, the marginal cost of transporting additional packets is essentially zero when the network has spare capacity. Thus a pricing mechanism is only required when congestion occurs [3].

A. Smart Market

Smart market [9] is a type of *congestion pricing* scheme. Each packet has a "bid" field in its header to indicate how much its sender is willing to pay for sending it. The packet will be admitted if the bid exceeds the current marginal cost of transportation in each router. Users do not pay the price actually bid, but only the market-clearing price, which is the bid of the lowest-priority admitted packet. This second-price auction scheme helps to maximize the ISP's revenue.

The value of each transferred packet is maximized and Internet users have an incentive to bid what they truly believe to be the value of transferring their packets. Although this mechanism has been considered to achieve both optimal distribution of capacity among users and network efficiency, it guarantees only a relative priority rather than an absolute QoS. A packet with a high bid gains access sooner than one with a lower bid, but delivery time cannot be guaranteed.

B. Shadow Pricing

The *shadow pricing* scheme (proportional fair pricing) [10] is proposed and applied to model a network in which a resource has the capacity to cope with a given number of equally sized packets in each time slot. Each packet arriving in overloaded slots is marked and charged a small fixed amount

(called a "shadow price"), and the mark is sent to the users. This shadow price will increase during times of congestion, and reduce during non-congested periods. The users then adjust their traffic load based on this feedback. Motivation for this approach is that resource characteristics—such as congestion information—are made known to the end users, who then determine themselves what should be their demands upon the network.

C. Edge Pricing

As contended by Shenker et al. [11], true *congestion pricing* is complex since the computations require knowledge of utilization not only for the user being charged but also of all other users who might be affected by the extra traffic—that is, the congestion conditions along the entire path. In reality, packets might take different paths depending on the condition of the network, and it is unfair to charge different users different amounts because of routing decisions that are beyond their control [3]. Thus, in *edge pricing*, the price is charged based on the expected (rather than instantaneous) congestion—depending on time of day, short-term congestion history, and so on—along the expected path appropriate for the packet's source and destination. The price therefore can be determined and charged at the access point, i.e., the edge of the provider's network, rather than computed in a distributed fashion along the entire path.

D. Congestion Discount

A *congestion discount* was proposed by Keon and Anandalingam [12] so that price would act as an incentive to shift traffic from congested periods to non-peak periods and, hence, balance the load. Customers are given choices of accepting a congestion discount rate and returning during a subsequent non-peak period, or rejecting the discount offer and obtaining the services immediately at a higher price.

E. Zone-Based Cost Sharing

In the *zone-based cost sharing* model, Clark [13] addresses the problem of sharing payment between senders and receivers. It is unfair for a payment structure in which "senders pay all" or "receivers pay all" since the senders and receivers, depending on the service, might wish to share the overall costs. Clarke proposed an additional field in the IP header to indicate a 'willingness to pay.' Its value indicates whether the sender, receiver, or neither of them is willing to pay for better than *best-effort* quality of service. The Internet is divided into different regions (zones), in which service is provided at a uniform, distance-insensitive way. The users can then specify the zones for which they are willing to pay.

F. Paris Metro Pricing

Odlyzko proposed the *Paris Metro pricing* model [14], inspired by the old city metro system of Paris, France. The old Paris Metro system offered two classes of cars, each identical in terms of both seats and arrival time, but the first class car would be charged twice as much as the second class. The idea was that the market would be self-regulating: first class customers were willing to pay more since they knew that the first class cars would be less crowded. Odlyzko suggests this model could be applied in Internet pricing by splitting the network into different channels, each with a fixed fraction of the network capacity. These channels treat packets exactly the same, the only difference being the price charged for usage. In principle, this scheme would result in higher QoS for users choosing the higher-priced channels.

Dube et al. [15] proposed something similar in the *Tirupati pricing* scheme, in which users pay different prices for joining different queues that are being served by the same server in a round-robin fashion. Significantly, the *Tirupati* scheme allows for the possibility of allocating more resources to one queue than to others.

G. Priority Pricing

Cocchi et al. introduced *priority pricing* [16] to enable multiple service levels over *best-effort* networks. Users enable priority flags for their traffic such as “service priority” or “no-drop.” If two packets arrive at the router simultaneously, the higher priority packet will be processed first—in the case of congestion, higher priority packets will have less delay. However, as the price is fixed, users might be paying more for priority service even when the network has spare capacity. Users might also end up receiving *best-effort* service when paying for priority service if the priority class is congested at a given instant in time. Gupta et al. [17] present a more sophisticated model where prices are updated at some time interval, according to the load and the congestion level of the network.

2. Pricing with QoS Guarantees

The integrated services (IntServ) architecture [18] was the first major attempt to enhance the Internet with QoS capabilities. It developed a new architecture for resource allocation to meet the requirements of real-time applications while preserving the datagram model of IP-based networks. The basic approach is per-flow resource reservation. The resource reservation protocol (RSVP) [19] has been developed as an end-to-end resource reservation setup protocol that installs a reservation state inside the network [20].

Unlike IntServ, the differentiated services (DiffServ) [21] architecture does not provide a complete solution for end-to-end QoS setup or management. DiffServ defines only a set of per-hop building blocks and a language with which to express per-hop forwarding behaviors. The research community has done additional work to build on top of the basic DiffServ framework.

The following sections introduce some major Internet pricing models developed for the IntServ and DiffServ architectures.

A. Charging for Integrated Services Using RSVP

Karsten et al. [22] propose a charging model that can be embedded in the RSVP architecture for an IntServ network. The basic idea is to use the PATH and RESV messages of RSVP to transmit pricing information and build a contract among senders, receivers, and the network. In addition to the usual traffic profile, the PATH message now carries a field with price information including the sender’s willingness to pay, the maximum share of costs, and the duration of price validity. At each hop of an outgoing link, the current market price for the requested QoS of the sender’s traffic is added to the price field by an ISP on the selected path. The price is determined by the ISP’s local pricing scheme, valid for a dedicated section of the end-to-end connection only.

It is possible for the final price to vary due to changes in the market situation. If the receivers agree with the pricing information provided in a PATH message, a RESV message is sent back to reserve resources. It also contains the calculated price to the sender. After this round-trip time the reservation and data transfer phase may proceed as usual with RSVP.

There are a number of critical issues of this proposed mode, including the time gap between the advertisement of available QoS and the time when a reservation is actually installed on a link. More research on QoS-sensitive routing needs to occur to support QoS-guaranteed reservations on the least expensive data paths. In addition, there are open questions on how to flexibly represent prices and price variations for different requests within a single service class, strategies for distributing charges for merged RSVP reservations, and the dynamics of pricing, payment methods, and security.

B. DiffServ Bandwidth Brokers as Mini-Markets

Fankhauser and Plattner [23] proposed a negotiation method between ISPs using a form of an advanced bandwidth broker, called a service level agreement (SLA) trader. Trading SLAs is performed between an ISP and its neighbors. Basically, ISPs offer their peers network resources that consist partly of resources directly owned by the network provider and partly of

resources bought by this ISP itself from other providers. On selling a service, ISPs charge other providers for the service through their network including the outgoing link from its egress nodes to the next autonomous system (AS). In this model, centralized SLA traders situated at each AS perform SLA trading. SLA traders make local decisions about what services are provided to which peers in a medium time scale (from several minutes to hours).

C. Resource Negotiation and Pricing Protocol

Wang and Schulzrinne [24] proposed a resource negotiation and pricing (RNAP) framework in which customers are able to negotiate and contract with the service provider for several QoS parameters such as peak rate, loss rate, and maximum delay. Both centralized and distributed ways of implementing RNAP are introduced in this model.

RNAP maintains negotiation intervals to manage the resource reservation contract. The negotiation interval defines the duration over which the negotiated service and price are valid. The contracted service expires automatically at the end of the negotiation interval, and the host's agent must periodically re-negotiate before the contract's expiration in order to ensure uninterrupted service. After re-negotiating the contract, the current service can be maintained with the previously negotiated price, with an updated price as required, or continue with best effort service.

The total charge to a customer is composed of three parts: a holding charge, usage charge, and congestion charge. Customers incur a holding charge if they do not fully use the contracted resources (for the resources they have not used) calculated at the price of a lower level of service than the contracted level. A usage charge is calculated based on the volume of the user's traffic and the price per traffic volume unit during the charging period. A congestion charge is made to the customers when their traffic travels through congested links of the network. The customer pays an extra congestion price for each of his/her packets traversing a congested area. The congestion price varies dynamically over time according to the congestion level of the particular link of the network.

D. Pricing over Congestion Control

Problems with pricing time scales are addressed by Yuksel et al. [25] in the *pricing over congestion control* (POCC) pricing model. In essence, POCC overlays congestion pricing on top of an underlying congestion control scheme, so that congestion in the interior network is controlled tightly at a small time scale, while pricing is done at a larger time scale compatible with human decision making in the loop. This model assumes that there is an underlying congestion control scheme, and that the

provider can set the parameters such that it leads to fairness and better congestion control. The pricing scheme on top can determine a user's willingness to pay and set the parameters of the underlying congestion scheme accordingly. Though the flows with a higher budget are treated more favorably, the overall system performance (e.g., fairness, utilization, and throughput) will be dependent on the flexibility of the underlying congestion control mechanism.

E. Charging for Actual Preferential Service Delivered

O'Donnell and Sethu recently proposed a pricing model for DiffServ [26] which charges users based on the bandwidth consumed, the instantaneous demand for the bandwidth, and the preferential service actually received by the user's traffic in the allocation of both bandwidth and buffer resources.

User traffic is divided into distinct classes of service. Let Q be the total number of classes of service, labeled $q \in \{1, 2, \dots, Q\}$ in order of increasing priority. Before sending a packet, the sender specifies which class of service that the packet should receive along the path (value of q). Routers queue packets separately based on the class of service to which the packet belongs. When a router's buffer space is full and a packet arrives from a service class q , a necessary number of packets from the service classes below q are dropped to make room for the new packet. Packets from the lowest class are dropped first. If there are no packets belonging to a lower service class, the new packet is dropped instead.

For billing purposes, every data packet carries a field in its control header representing the price that will be charged for the packet's transmission along the path. Each router in the path adds its charge to this field in the packet (based on three separate components: bandwidth consumed, pricing preferential service rendered, and buffer resources occupied).

When the packet reaches its destination host, this price field is copied into its acknowledgement and returned back to the destination host. Based on the recent history of prices charged, the user may dynamically adapt and adjust its sending rate or select a different class of service.

O'Donnell and Sethu's pricing scheme was based on the framework of *shadow pricing*. However, their scheme does not assume a prior contract between the network and the customer for a certain QoS. Instead, they rely on the user adapting their transmission rate and/or the service class based on the price feedback it receives for each packet. Users are only charged for the QoS they've actually received.

IV. A Three-Dimensional Evaluation Model

A viable Internet pricing scheme needs the support of both economic tools (including accounting, charging, billing, and

pricing strategies in resource allocation) and technological mechanisms (for example, congestion control, QoS technologies, user authentication, and system security). Figure 2 shows broadly how Internet pricing research stands in relation to supported research areas.

Implementation effort is also of great importance when assessing the practicality of any proposed pricing schemes. We propose examining pricing schemes against a 3D metric that captures the contribution (dimensions) of technical efficiency, economic efficiency, and social impact.

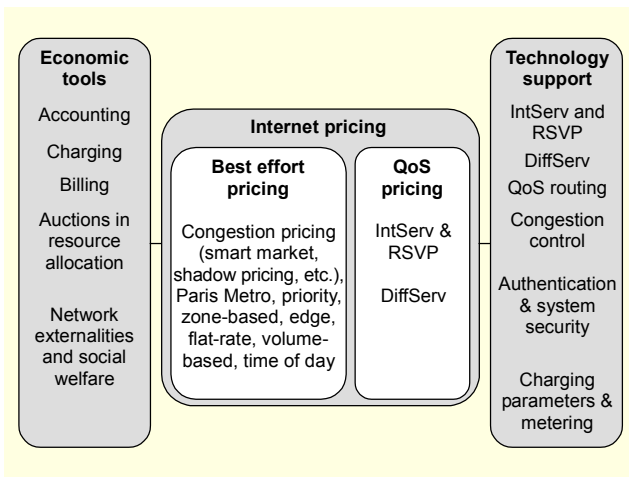


Fig. 2. Scope of Internet pricing research.

1. Defining the New Model

Technical efficiency¹⁾ refers to the technical costs associated with applying the new technology of a particular pricing model or scheme—e.g., upgrading existing equipment, the overhead of the financial system (accounting, charging, and billing), and labor (the cost of training and employing qualified technical personnel in order to operate the newly upgraded equipment and software). A pricing model is considered more ‘technically efficient’ if it can reduce these costs.

Economic efficiency captures the impact of a pricing scheme on network utilisation and the optimization of a service provider’s revenue. This dimension reflects the ability to handle additional customers without upgrading links, the possibility of attracting new customers due to cheaper traffic options and/or improved QoS, the capability of accommodating new Internet services and valued customers, and the maximization of marginal costs in charging customers’ traffic.

Social impact concerns the fairness among network users.

1) In [27] we used the term “technical complexity”, which referred to the implementation cost of a pricing model. However, in visualizing the evaluation model, we would like to keep the consistency of the value interpretation of all three dimensions – the higher the value the better – so we changed it into “technical efficiency”

Users with more valuable traffic will be given more network resources and better quality of services if they have greater willingness to pay than others. Also, the price these users have to pay takes into account the marginal social cost of the extra congestion that their traffic creates for others.

For most pricing schemes there is a distinct coupling and inter-relationship between economic efficiency, social impact and technical efficiency. For example, a more finely grained charging unit and dynamic pricing system will lead to a fairer allocation between users and a better maximization of both the service providers’ revenue and network utility. However, these come at a cost of technical efficiency. More finely grained charging units lead to higher accounting overheads and processing costs (per charging unit). A highly dynamic scheme (reacting faster to network status such as network congestion) comes with the higher cost of internal communication overheads between senders, network nodes, and receivers.

Clearly, it would be nice to discover an optimal pricing model where economic efficiency, social impact, and technical efficiency are all concurrently maximized. However, practical pricing models will always reveal a trade-off between the three dimensions in our evaluation scheme. Figure 3 provides a qualitative illustration of how an optimal pricing model and a practical pricing model might visually differ in our model (in this example, each of the dimensions has an arbitrary range of 0 to 10, where 10 is the most desirable and 0 is the least desirable).

It might be argued that an Internet pricing model is inherently multi-dimensional, with many more variables than the three axes we have described here. While this is true in detail, the value of our approach is to reflect the many discrete variables into three largely orthogonal, aggregate variables. The 3D model summarises key metrics in a way that simplifies comparisons between Internet pricing schemes.

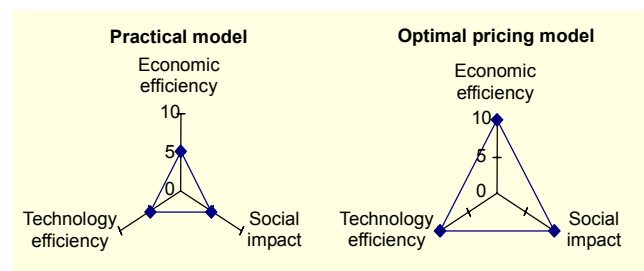


Fig. 3. Practical pricing model vs. optimal pricing model.

2. Using the 3D Model to Evaluate Pricing Schemes

We can demonstrate the qualitative value of our 3D metric when comparing, for example, *smart-market pricing* and *flat-rate pricing*. Figure 4 shows that these two schemes have distinctly different appearances when mapped into our 3D metric.

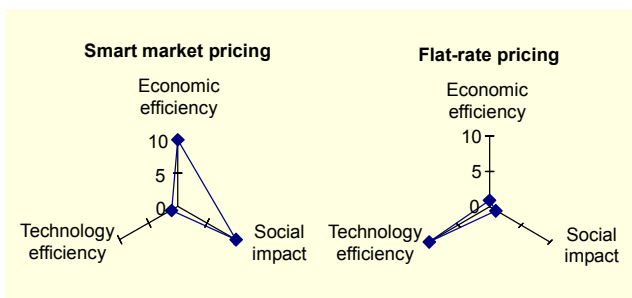


Fig. 4. Smart-market pricing model vs. flat-rate pricing model.

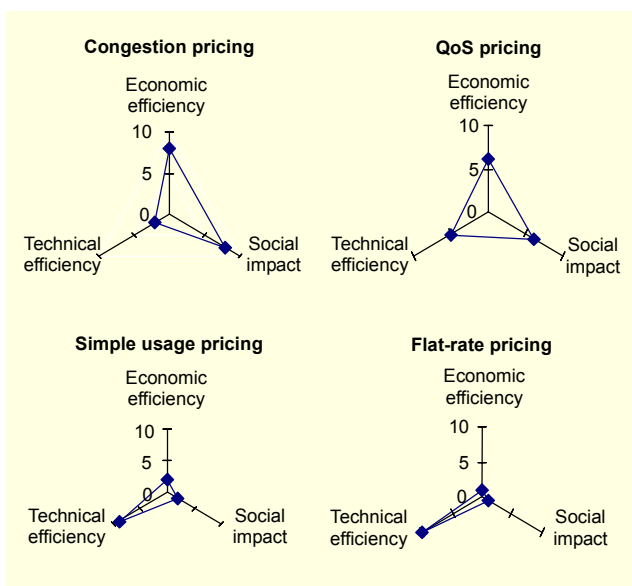


Fig. 5. Technical and economic efficiency, and social impact.

Congestion pricing maximizes the economic efficiency and social welfare (charging at the packet-level and taking into account the social cost of delivering a packet). *Smart market* is an extreme of this pricing category. By charging per packet and deploying the second-bid auction scheme, it maximizes the revenue for the ISPs. However, it has many implementation problems. It requires interior routers to sort packets according to their bids, thus requiring the deployment of new routers. Extra packet header fields are also required, and there is the additional communication overhead of the auctioning process to be considered. By contrast, flat-rate pricing incurs no special implementation costs but provides very low economic efficiency and social impact.

For most proposed pricing models with QoS guarantees, the economic efficiency and social impact goals are achieved to some extent, varying from an absolute QoS guarantee to a different relative priority. However, technical efficiency remains a concern—many QoS guarantee models are constrained by their requirement to upgrade all routers to support both the per-hop QoS mechanisms and the charging system.

Self-regulating schemes such as *Paris Metro pricing* and *congestion discount* might not work under competition with the nature of the current Internet. Service providers might lower their prices in response to competition, and economic efficiency and social-impact dimensions are hardly achieved. Table 1 provides a comprehensive comparison of many pricing models based on the dimensions of our 3D metric. Figure 5 summarizes the relative weights of our three dimensions for four different pricing model categories²⁾.

V. Factors Inhibiting Deployment of Proposed Pricing Schemes

Implementation cost is one of the major factors hindering deployment of many of the proposed pricing schemes. This section discusses a number of other important inhibiting factors and returns to the implementation factor in a cost benefit analysis from an ISP's point of view.

1. The Consumer Perspective

First, we examine the schemes in section III from a consumer perspective. Both the INDEX [29] and CATI [30] projects show that the most important requirements and expectations from the users' perspective are transparency and predictivity of a pricing scheme—providing the users' detailed charging information (such as per flow of traffic, per sessions, or per different services) and enabling users to predict or estimate the costs of using the network service [22], [31]. Charging schemes that adapt according to an internal network state rarely meet this criteria (excepting the highest bidding values in auction-based charging schemes). Prices can be updated in a short time-scale if the user-provider negotiation is done automatically, while negotiation requiring human intervention might prefer a longer and more stable time scale.

Another critical user requirement is stability—the assurance of QoS provided to users [22], [31]. In *congestion pricing*, users only pay more to gain a higher priority for their traffic, they do not gain any QoS assurance. Pricing for IntServ using RSVP provides hard QoS guarantees. However, users would be irritated when dynamic price changes result in the tear-down of their reservation mid-session. In terms of charging parameters (such as delay, jitter, and loss) we also need an exact definition of when “quality assurance is met.” Users must be able to estimate the impact of such quality

²⁾ The value for each dimension presented in the figure is speculative rather than experimental in nature. We just present them here to demonstrate the visual effect of our evaluation model. The mapping between qualitative assessments to numerical representations for each dimension is a challenge and will be covered in our future work.

Table 1. Summary of the reviewed models.

Pricing scheme	Technical dimension			Economic dimension		Social welfare dimension	
	Require upgrade equipment	Accounting overhead	Communication overhead/ packet overhead	Charging method	Charging unit	Service levels and QoS guarantee	Senders & receiver sharing cost
Smart market	All routers along the path	Very high	Very high	Second-bid auction	Per packet	Only relative priority	No
Shadow pricing	All routers along the path	High	No	Dynamic congestion pricing	Per packet	No QoS guarantee	No
Edge pricing	Edge routers	Low	No	Congestion pricing	Per contract	No QoS guarantee	No
Congestion discount	No	Low	No	Congestion pricing	Per packet	No QoS guarantee	No
Zone-based cost sharing	All routers along the path	High	No	Split charge between sender & receiver	Per packet	Not clear	Yes
Paris metro	All routers along the path	High	No	Static charge per class of service	Per packet	No QoS guarantee	No
Priority pricing	All routers along the path	High	No	Static charge per class of service	Per packet	Only relative priority	No
IntServ & RSVP	All routers along the path	High	Very high	Auctioning for resource reservation	Per flow	Absolute QoS	Yes
RNAP	All routers along the path	High	Very high	Negotiation by contract	Per contract	Negotiated by contract	No
Actual service received	All routers along the path	High	High	Charge only for actual QoS received	Per packet	No QoS guarantee	No

goals on their applications and see evidence that QoS targets have been met.

Pricing schemes should also be flexible and user friendly. Zone-based cost sharing and IntServ pricing with RSVP provides the ability to share costs between senders and receivers. However, IntServ pricing with RSVP is inflexible when it forces receivers to compete for resources along a common, and potentially congested, shortest path back to the source. The ability to switch between competitive ISPs remains an open issue.

2. The ISP's Perspective

Next, we look at the schemes from an ISP's perspective. Clearly, implementation costs are critical and must not exceed the revenues likely to be gained by deploying any new scheme. On the other hand, benefits come from improved network utilities and marginal revenues from valued customers. Figure 6 summarizes the related costs and benefits. An effective solution maximizes the consequences of implementation costs and benefits (of which additional revenue is just one).

Network stability and reliability must also be considered. ISPs resist deploying a complex technology if there are questions as to its reliability and operational effort. In this context, "throwing bandwidth at the problem" is attractively simple and reliable. Pricing schemes that require equipment

upgrades must work around the reality of incremental end-to-end upgrades; they must consider backward compatibility with older parts of the network and different service domains.

The settlement processes between ISPs are also not addressed well in most proposals. For example, DiffServ is claimed to solve the scaling problem of IntServ technology; however, it lacks standardized service classes between ISPs—making a settlement between ISPs more difficult. Fraud protection and legal security are also important in charging for end-to-end QoS. In a multi-ISP environment, in case of a failure, there should be enough information to determine who has the liability for the failure. Table 2 summarizes the reviewed models according to some of these inhibiting factors.

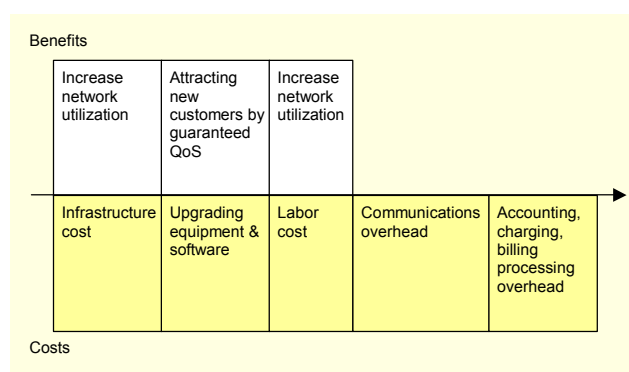


Fig. 6. Costs and benefits from an ISP's perspective.

Table 2. Summary of the reviewed models according to some inhibiting factors.

Pricing scheme	Transparency & predictability	Stability of service	Price time scale
Smart market	Unpredictable & non-transparent	Unstable & no guarantee	Price fluctuates per packet
Shadow pricing	Unpredictable	No guarantee	Price fluctuates depending on the network's condition
Edge pricing	Predictable to some extent	No guarantee	Dynamic pricing depends on network condition and estimation
Congestion discount	Predictable	No guarantee	Dynamic pricing depends on network condition and estimation
Zone-based cost sharing	Predictable	No guarantee	Price fluctuates per packet
Paris metro	Predictable	No guarantee	Static over long time scales
Priority pricing	Depends on the actual implementation	No guarantee	Depends on the actual implementation
IntServ & RSVP	Unpredictable	No guarantee during re-negotiation period	Dynamic per contract interval
RNAP	Predictable	No guarantee during re-negotiation period	Dynamic per negotiation interval
Actual service received	Unpredictable	No guarantee	Depends on the network's condition

VI. Possible Pricing Solutions for the Internet

We suggest short-term and long-term Internet pricing solutions, reflecting the constraints of supporting QoS mechanisms, time scales for technology standardization, and the current Internet context. Currently, there is a limited deployment of QoS technologies (such as IntServ and DiffServ) which discourages research into QoS pricing schemes. On the other hand, a lack of supporting pricing schemes inhibits deployment of new QoS technologies. Therefore, the most appropriate short-term solutions would be pricing models built around *best-effort* traffic and mechanisms. However, in the long-term, additional demand for a 'hard' QoS will make efficient QoS pricing schemes more desirable.

For example, a short-term Internet pricing solution would combine flat-rate pricing, usage pricing, and *congestion pricing* with a compromise of implementation costs and benefits. Flat-rate pricing covers the fixed costs of services while usage and *congestion pricing* controls congestion, differentiates service by different charging levels, increases social welfare and fairness among Internet users, and produces improved marginal revenues for service providers.

Long-term Internet pricing schemes must allow a predictable establishment of QoS tied closely with measurable charging parameters (such as bandwidth, delay, jitter, and loss) and unambiguous overall financial consequences for users. There are no clear solutions at this stage that account for user control of routing, cost sharing between senders and receivers, and standardized settlements between ISPs. Authentication and legal security issues also need to be solved in the longer term.

VII. Conclusions

This paper has presented a novel 3D model for qualitatively analyzing and evaluating Internet pricing schemes based on orthogonal metrics of technical efficiency, economic efficiency, and social impact. We have reviewed Internet pricing research over the last ten years and compared and evaluated the reviewed models using our 3D visualization model. We have also highlighted the possible factors that have inhibited the deployment of the discussed models and proposed short-term and long-term Internet pricing solutions.

A viable pricing scheme will be a trade-off between technical efficiency, economic efficiency, and social impact. For example, a simple, low-cost and easy-to-explain scheme with moderate economic efficiency might be preferable to a scheme showing optimal economic efficiency but complex and costly implementation.

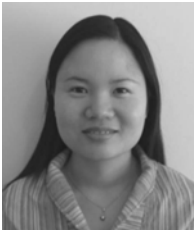
Most of the models reviewed here have been theoretical or speculative rather than experimental in nature, so it is difficult to make clear and strong assessments to their worth. Although our evaluation model based on a 3D metric addresses critical aspects of a pricing model, it needs to be supported by a strong experimental base. Future research, therefore, should include a detailed evaluation of the technology implemented for the proposed pricing schemes.

Acknowledgment

We would like to thank the anonymous reviewers for their valuable feedback to improve this paper.

References

- [1] "Who pays for the Internet?," http://mmm.wiwi.hu-berlin.de/IMI/s_org-provider.html, 1997.
- [2] R. Ruth, Usage Accounting for the Internet, http://www.isoc.org/isoc/whatis/conferences/inet/97/proceedings/F1/F1_1.HTM.
- [3] T. Henderson, "Internet Economics – A Literature Review," <http://www.cs.ucl.ac.uk/staff/T.Henderson/doc/litreview.ps.gz> (accessed Apr. 2000).
- [4] B. Stiller, G. Fankhauser, N. Weiler, and B. Plattner, "Charging and Accounting for Integrated Internet Services – State of the Art, Problems and Trends," *Proc. INET '98*, Geneva, Switzerland, July 1998.
- [5] R. Cocchi, S. Shenker, D. Estrin, and L. Zhang, "Pricing in Computer Networks: Motivation, Formulation and Example," *IEEE/ACM Trans. on Networking*, vol. 1, no. 6, Dec. 1993.
- [6] S. Shenker, "Some Fundamental Design Decisions for the Future Internet," *IEEE J. on Selected Areas in Comm.*, vol. 13, no. 7, 1995, pp. 1176-1188.
- [7] G. Fankhauser, B. Stiller, C. Vogtli, and B. Plattner, "Reservation-Based Charging in an Integrated Services Network," *Proc. 4th INFORMS Telecommunications Conf.*, Boca Raton, Florida, USA, Mar. 1998.
- [8] M. Karsten, J. Schmitt, L. Wolf, and R. Steinmetz, "An Embedded Charging Approach for RSVP," *Int'l Workshop on Quality of Service '98*, Napa, California, USA, May 1998.
- [9] MacKie-Mason K. J. and Varian R. H., "Pricing the Internet," Memorandum 20/1993, Oslo University, Dept. of Economics, Apr. 1993.
- [10] F.P. Kelly, "Charging and Accounting for Bursty Connections," *MIT Workshop on Internet Economics*, Mar. 1995.
- [11] S. Shenker, D. Clark, D. Estrin, and S. Herzog, "Pricing in Computer Networks: Reshaping the Research Agenda," *ACM Computer Comm. Review*, vol. 26, 1996, pp. 19-43.
- [12] N. Keon and G. Anandalingam, "A New Pricing Model for Competitive Telecommunications Services Using Congestion Discounts," <http://mail3.rhsmith.umd.edu/Faculty/KM/papers.nsf/0d5ea3f525a84fc5485256d0c006f210d?OpenDocument> (accessed July 2000).
- [13] D. Clark, "Combining Sender and Receiver Payments in the Internet," <http://www.gta.ufrj.br/DiffServ/csrp-ddc.ps.gz>.
- [14] M. Odlyzko, "Paris Metro Pricing for the Internet," *Proc. 2nd Int'l Conf. on Information and Computation Economics, (ICE099)*, Nov. 1999.
- [15] P. Dube, V. Borkar, and D. Manjunath, "Differential Join Prices for Parallel Queues: Social Optimally, Dynamic Pricing Algorithms and Application to Internet Pricing," *Proc. INFOCOM 2002*.
- [16] R. Cocchi, D. Estrin, S. Shenker, and L. Zhang, "A Study of Priority Pricing in Multiple Service Class Networks," *Proc. SIGCOMM Symp. on Comm. Architectures and Protocols*, Switzerland, Sept. 1992.
- [17] A. Gupta, D. Stahl, and A. Whinston, "Priority Pricing of Integrated Services Networks," http://cism.bus.utexas.edu/alok/joc_paper/mit_paper.html.
- [18] RFC 1633, *Integrated Services in the Internet Architecture: An Overview*, IETF, 1994.
- [19] RFC 2205, *Resource Reservation Protocol (RSVP) – Version 1 Functional Specification*, IETF, 1997.
- [20] Z. Wang, *Internet QoS Architectures and Mechanisms for Quality of Service*, Morgan Kaufmann Publishers, San Francisco, 2001.
- [21] RFC 2475, *An Architecture for Differentiated Services*, IETF, 1998.
- [22] M. Karsten, J. Schmitt, L. Wolf, and R. Steinmetz, "Cost and Price Calculation for Internet Integrated Services," *Proc. KiVS'99*, Darmstadt, Mar. 1999.
- [23] G. Fankhauser and B. Plattner, "DiffServ Bandwidth Brokers as Mini-Markets," <http://www.tik.ee.ethz.ch/~cati/paper/isqe99b.pdf>.
- [24] Wang and H. Schulzrinne, "A Framework for Resource Negotiation and Pricing in the Internet," <http://www1.cs.columbia.edu/~xinwang/public/paper/RNAP.pdf>.
- [25] M. Yuksel, S. Kalyanaraman, and A. Goel, "Congestion Pricing Overlaid on Edge-to-Edge Congestion Control," *Proc. ICC 2003*, Anchorage, AK, May 2003.
- [26] J. O'Donnell and H. Sethu, "Congestion Control, Differentiated Services, and Efficient Capacity Management through a Novel Pricing Strategy," *Computer Comm.*, vol. 26, issue 13, Aug. 2003, pp. 1457-1469.
- [27] T. Nguyen and G. Armitage, "Pricing the Internet a Visual 3-D Evaluation Method," *Proc. ATNAC 2003*, Melbourne, Australia, Dec. 2003.
- [28] J. Almann, B. Rupp, and P.P. Varaiya, "Effects of Pricing on Internet User Behavior," *NETNOMICS*, vol. 3, no. 1, June 2001, pp. 67-84.
- [29] R.J. Edell and P.P. Varaiya, "Providing Internet Access: What We Learn from the INDEX Trial," *Technical Report*, University of California at Berkeley, CA, Apr. 1999.
- [30] Charging and Accounting Technology for the Internet (CATI) Project, <http://www.tik.ee.ethz.ch/~cati/>
- [31] B. Stiller, P. Reichl, and S. Leinen, "Pricing and Cost Recovery for Internet Services: Practical Review, Classification and Application of Relevant Models," *NETNOMICS - Economic Research and Electronic Networking*, Baltzer, The Netherlands, vol. 3, no. 1, Mar. 2001.
- [32] P. Key, D. McAuley, and P. Barham, "Congestion Pricing for Congestion Avoidance, Microsoft Research," *Technical Report*, Cambridge, UK, Feb. 1999.



Thuy T.T. Nguyen is a PhD candidate at the Centre for Advanced Internet Architectures, Swinburne University of Technology, Melbourne, Australia. She received the B. Eng. Dip. Prac. in telecommunications engineering from the University of Technology, Sydney in 2002. Her research interests include Internet pricing and charging systems, traffic characterization and measurement, QoS, and the performance evaluation of broadband and wireless networks. She is a member of IEEE Communications Society.



Grenville J. Armitage is an Associate Professor of telecommunications engineering and Director of the Centre for Advanced Internet Architectures at Swinburne University of Technology, Melbourne, Australia. His research interests include networked games, IP traffic pattern analysis, broadband IP access architectures, and network security. He has a PhD in electronic engineering from the University of Melbourne, Australia. From 1994 to 1997 he was a research scientist with Bellcore's Applied Research Division (now Telcordia Technologies) and from 1997 to 2001 held research positions with Bell Labs Research (New Jersey) and Bell Labs Research Silicon Valley (California).