ISPs - Pricing Internet Access

Abstract
Access to the Internet is provided by a number of commercial entities known as Internet Service Providers who constitute the Internet backbone and act as mediators between the user and the Internet. A variety of pricing methods have been considered in the literature and implemented in practice. While congestion control is necessary condition for the smooth operation of the global Internet and affects all users, demand of service differentiation is related to the requirements of specific applications using the Internet as an infrastructure. The pricing model of an ISP would have to be able to accommodate the levels of service offered.

Track 10, keywords: ISP, Internet access, pricing models

1. Introduction
The physical topology of the Internet provides a set of limited resources – namely point-to-point bandwidth and buffer storage. Internet users contesting the use of these resources are humans, or applications and/or processes (Key, 1999). The Internet - a “network of networks”, occupies the middle layer of the (Open Systems Interconnectivity) model shown in Figure 1. It must be noted that while the contested resources are supplied within the lower layer (where the infrastructure is provided by national and international telecommunications services providers), resource allocation can occur across all layers.

Access to the Internet is provided by a number of commercial entities known as ISPs (Internet Service Providers), who constitute the “Internet backbone”. All ISPs offer at least one basic service – Internet traffic routing. In Figure 2 we show a hierarchical view of the Internet backbone, which comprises two types of ISPs: “smaller” ones who offer services to individual users and business but have to purchase connectivity options, and “large” ISPs who offer services directly to users and sell some of their resources to the smaller ISPs. A New Zealand example of an ISP of the first type is “Ihug” – a company which leases some of
its telecommunications infrastructure from Telecom New Zealand, who in turn act as a large ISP through a subsidiary (Telecom’s “Xtra”).

The role of an ISP is that of mediator between the user and the Internet (Greenstein, 2001). Two important characteristics of the services provided by ISPs are highlighted: firstly, there seems to be an “inherent uncertainty” over the economic value of their service, and secondly, both the cost and the value of the technological mediation vary “across users and over time”. Trying to solve the economic problems arising from these two specific features of paid Internet access provision, a variety of pricing methods have been considered in the literature and implemented in practice.

![Figure 1. The Internet – a layered representation](image)

Two factors contribute significantly to the problems identified – the connectionless nature of Internet routing, the lack of a working mechanism for providing differentiated application layer services across the Internet as a whole. As shown in Figure 2, a “paid for” packet might be part of a whole message M travelling from User 1 to User 2 across the Internet - with different packets taking different routes. While the cost of a packet A send from User 1 to User 2 should incorporate the costs incurred on all legs of the path, the legs might belong to different ISPs.

In comparison to the connectionless network established by the Internet protocol (IP), the underlying infrastructure (bottom layer in Figure 1) is connection oriented. A connection-oriented telephone system, for example, establishes a permanent or a semipermanent path for each call and reserves the resources necessary to support the call. In
contrast, a connectionless IP packet travels from one IP router to another without “knowing” whether or not there is a path ahead. IP packets belonging to the same message may travel through different routers, and “meet” only at the receiving end. In the connection-oriented telecommunications network the “sender” sets a call and makes sure that the “receiver” is accessible and the path is available (otherwise the sender gets the busy signal). In the connectionless IP network there is never a busy IP signal, and decisions are made by all routers on the path of an IP packet. The state of “connectionless-ness” might lead to network congestion – it would occur when IP “senders” attached to the network send traffic in excess of the network capacity. As a result some IP packets will be lost. The transport layer protocol (TCP) uses the number of packets lost as a feedback mechanism to determine the rate at which to send IP traffic and thus to introduce a certain level of control.

Figure 2. The Internet – IP routing

Telecommunication networks bill customers either on usage, or using a flat fee mechanism for some types of calls (for example, local calls in New Zealand are covered by a monthly subscription fee). Even in the case of flat fee pricing, a telecommunication network
can control congestion by simply refusing to set up calls when resources are not available. In contrast, Internet pricing using a flat fee model (or even a traffic based fee) provides no mechanism for network congestion control.

While congestion control is necessary condition for the smooth operation of the global Internet and affects all users, demand of service differentiation is related to the requirements of specific applications using the Internet as an infrastructure. The current practice in providing differentiated services is rapidly evolving towards a quality of service (QoS) paradigm, where ISPs will be able to offer users serviced based on low jitter (eg for video streaming), or low delay (for voice over IP). The pricing model of an ISP would have to be able to accommodate the levels of service offered. The path between a supplier of an application service and a user would be covered by number of ISPs - who currently do not settle among themselves. An ISP charges its own customers and retains the full amount charged, from which to recover costs and obtain profit. This aspect of the global structure of Internet access provision also affects the choice of the ISP’s pricing model (Brownlee, 1994; Crawford, 1996; Friedman & Mills-Scofield, 1998).

![Diagram of the value chain of Internet access provision](image)

**Figure 3. The value Chain of Internet access provision**

The rest of the paper is organised as follows: the next section provides a classification of Internet access pricing models and discusses a selection of usage-sensitive pricing models. The last section compares the relative advantages and disadvantages of the three arguably simplest differential pricing models, and flat rate pricing. The value chain of Internet access provision is shown on Figure 3 (adapted from Dolan, 2000) is used to identify the dependencies between the three types of ISPs providing services to end users and applications, and to consider the possible adoption of a model in the context of the ISP position in the chain. Usage-Sensitive Pricing Models
Among the methods used to price user access to the Internet, by far the most popular is the flat pricing one. In this model the user pays a flat recurring fee for a fixed period of time (e.g. monthly). As pointed out in (McKnight & Bailey, 1996; Wiseman, 2000) flat pricing is convenient for accounting purposes and reduces administrative overhead. However, flat pricing does not address two major issues in Internet usage: a) the need to control (and possibly reduce) Internet traffic congestion, and b) the need to prioritise access for QoS applications. Alternative pricing models suggesting solutions to the problems stated have been suggested: usage based models, the auction approach, static and dynamic priority based models, the “Paris Metro pricing” model, differential pricing models (Crémer, 1999; Dolan, 2000). A classification of the basic Internet access pricing models is presented in Table 1, followed by a more detailed discussion of a selection of non-flat rate models.

Table 1. A basic classification of Internet pricing

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flat-rate pricing</td>
<td>A fee is paid to connect. Individual bits sent/received are not priced. The user does not pay for any specific quality. The user pays regardless of the topology.</td>
</tr>
<tr>
<td>Non flat-rate models</td>
<td>Usage-sensitive pricing: A portion of the fee is for the connection. The second portion of the fee is related to the volume (i.e. each bit sent or received has a marginal non-zero cost).</td>
</tr>
<tr>
<td>(usage-based)</td>
<td>Transaction-based pricing: Every bit sent or received has a marginal non-zero cost, determined not by volume but by the characteristics of the transaction.</td>
</tr>
</tbody>
</table>

2.1 Pricing models based on the IP protocol

The telephone-pricing model (MacKie-Mason, 1993) is based on the notion of “distance” measured by the hop count as used by IP routers. The proposed model has two major faults: firstly, the TCP/IP protocols do not “minimise distance” but control congestion through choosing alternative routes, and secondly - posted prices are not flexible enough to let users who are willing to pay more use the resource with a priority. To handle the second issue, the "precedence" model proposed in (Bohn et al, 1994) utilises the special “precedence” field in the IP packet header to assign priority to a packet. The proposal does not specify the mechanism for assigning priority to a packet, and the mechanism for updating an existing priority scheme.
2.2 The “smart market” model

The model was first proposed in MacKie & Varian (1993), and further developed and discussed in MacKie & Varian (1994; 1995; 1996). MacKie & Varian point out that a congested network introduces a social cost (deteriorating response time) and conclude that a “desirable pricing structure is one that allocates congested bandwidth and sends appropriate signals to users and network operators about the need for expansion and capacity”.

![Figure 4. “Smart market” supply and demand](image)

The “smart market” solution to the problem revolves around a mechanism of setting the price for network access at different priorities. Users indicate their “maximum willingness to pay” for network access, i.e. they “bid” for network access. The bid is attached to each data packet. Routers examine the bid field and admit packets with bids higher than a pre-determined cut-off value. The price $P^*$ charged to all users is determined by the intersection of the supply and demand curves (Figure 4, where the supply of bandwidth is a constant $K$). The model offers a mechanism for prioritising – packets with higher bids will be admitted before packets with lowed bids. It can be viewed as a Vickrey auction where $N$ highest bidders obtain network access at the $(N+1)^{st}$ highest price bid. The implementation of the mechanism would bring an accounting overhead and significant changes to routing algorithms but would be able to send the correct signals for capacity expansion “when the revenues from congestion fees exceed the cost of providing the capacity” (MacKie & Varian,
1994). The smart market model can be extended to provide QoS pricing (e.g., multimedia) but the computational demands would increase (MacKie & Varian, 1996).

2.3 The “expected capacity” model

The model was proposed by Clark (1996). Based on the hypothesis that the user criterion for delay in a congested network is not the delay of each packet as experienced by the network but rather the total elapsed time to transfer one typical application element. (Some examples are an average 2K Web page, or a single keystroke). Therefore, service quality is determined by the degree to which a user is dissatisfied if the target delay of the transaction is not met. An extreme case of dissatisfaction are real-time applications where a delayed packet might lose its utility – compared to “elastic” applications which are more tolerant to delays. Clark’s model introduces a pricing mechanism which can be used to control the allocation of service according to the needs of different users.

The expected capacity scheme proposes to allocate service among users in a way which offers a range of expectations rather than a range of guarantees. Both subscription pricing (fixed fee) and volume-based pricing do not meet the criterion of user satisfaction as defined by Clark as they impose charges regardless of the state of the network – congested or not. Clarke’s solution ties pricing to the expected capacity. It does not restrict the user’s ability to send packets if the network is not overloaded, while charging for packets with non-zero marginal sending costs.

As shown in Figure 5, Clark’s model is based on packet tagging - a concept implemented in wide area networking technologies such as Frame Relay and ATM (Asynchronous Transfer Mode). The user negotiates an expected capacity contract with the provider at the point of access. All packets sent are examined and tagged as “in” or “out” of the envelope of expected capacity. At routers, where there is no congestion, traffic control is not based on tags. Tagged packets are dropped (or explicit notification is sent back to the sender) when congestion occurs. The method ensures that a flow of packets which stays
within its expected capacity is less likely to be dropped compared to a flow which tried to go faster. While there is no provision of a guaranteed minimum capacity or a predefined service class, the model aims to provide best service at the macro- (application) level rather than on the micro- (packet) level. A settlement mechanism to be used by multiple Internet service providers is also suggested.

Figure 5. Tagging packets in Clarke’s model (Source: Dolan, 2000)

2.4 The “Paris Metro Pricing” (PMP) model

The model proposed by Odlyzko (1999a; 1999b) is based on a class separation approach implemented by the Paris metro. The pricing scheme model does not offer a service based on the estimated demand for quality. A number of fixed quality classes are stipulated and demand is regulated through feedback. As shown in Figure 6, the network is subdivided into logical network channels (classes) with fixed capacity and fixed usage price. Once the prices are set, users are left to make a decision about class they would use. Knowledge about application requirements, pricing (budget constraints) the current level of congestion in each class informs the user decision and determine the QoS to be received. The Internet is “subdivided” into different logical networks, which perform at a predictable QoS level.
Traffic management is reduced to asking users to differentiate their requirements through selecting different class channels. The method provides inexpensive congestion control and does not require major changes to the network. It is relatively simple and close to flat rate pricing and should be acceptable to users, who are perceived to be tolerant to substantial variations in the quality of service they receive provided that the pricing scheme is clear and predictable. To minimize losses incurred due to traffic segregation, the number of channels should be kept down (up to 4), thus satisfying consumer preferences for simplicity.

PMP could be able to work together with other application layer QoS protocols (such as RSVP- ReSerVation setup Protocol), which could operate in one or more classes. The model is compatible with some other proposed pricing schemes – for example, with the precedence model (Bohn et al, 1993) and with the priority model in (Gupta et al, 1997). While fixed capacity allocation would be suitable for the Internet core (the large ISPS in the value chain), routing decisions at the edge could be based on weighting. Different approaches could be implemented (in a single or a mixed mode) in different parts of the network. Odlyzko (1999a) gives the following example: “. one could assign 40% of the capacity of the network to class 1 traffic, and 60% to classes 2 and 3, with weighted priority queuing determining what packets in classes 2 and 3 are to be sent first...”. For predictable service, capacity allocation should remain stable for periods of time but day-night differentiation at the edge might encourage better utilization. PMP charges could comprise a combination of a
fixed charge per packet and a variable, size-dependent charge per packet – to encourage the
development of applications which generate small-size packets.

Several other approaches towards differential Internet access pricing are described
in the literature. One example is the pricing model based on the general equilibrium theory
(Gupta et al, 1997). It is based on stochastic equilibrium, in which average service rates are
optimal for each user, given the price and the anticipated delay; the anticipated delays are
the “correct ex-ante anticipated delays given the average flow rates”. The rental classes for
services are defined using a formula derived from the model. Special network nodes (servers)
offer the end-user choice of priority classes with associated price – with the price increasing
as traffic volumes through a particular server increases, to control congestion. The authors
have carried extensive simulations to examine the work of the proposed pricing algorithm

Other works in the area are include Crémer (1999), who proposes a stationary
model to approach to the pricing of network capacity, Fulp & Reeves (2001) – who develop a
framework of strategies for pricing and provisioning for a differentiated service network
consisting of hierarchical markets, Stiller, Gerke, Reichl & Flury (2001) – who consider
charging and pricing for differentiated services as an element of the extended functionality of
network management, and also Edell &Varaiya (1999) - who describe an alternative ISP
model with built-in functionality to offer differentiated QoS on demand and with prices
which reflect resource cost. Due to the constraints of this paper we would omit further but
will focus on the comparison of the three simple differential models and the flat rate pricing
and will illustrate the current trends with data from New Zealand.

3. Discussion and conclusion

The two earlier models mentioned in section 2.1 have become obsolete due to the
new developments in exterior IP routing as a well as the move towards IPv6. Considering
also (Dolan, 2001) and Mason (2001), we summarise the features of the other models
discussed above, as shown in Table 2. The methods are compared in terms of technical
implementation, user acceptance, guarantee of service, and capacity and congestion control. The expected capacity model is not complete as it does not include the need to install new software in the Internet core. User acceptance might be affected by the lack of a guarantee of service. Similarly, the smart market pricing scheme offers allocation of resources per packet, but not per an end-to-end connection. In addition, the complexity of the required router modification would contribute to rather than alleviate network congestion. The PMP model is relatively simple, and given the additional advantage of being compatible with other pricing schemes, is the most attractive of the three.

Table 2. A Comparison of three pricing models: “smart market”, “expected capacity” and “Paris Metro Pricing”

<table>
<thead>
<tr>
<th></th>
<th>Smart Market</th>
<th>Expected Capacity</th>
<th>Paris Metro Pricing</th>
</tr>
</thead>
</table>
| **Technical implementation** | • Some data link layer protocols can accommodate the bid tag; IPv4 is not suitable, possible IPv6.  
• Routers need modification to implement billing – to sample every Nth packet | • A token is generated independently of the packet.  
• Special software needed at routers or other nodes (not addressed in the proposal) | • The priority field in the IPv4 header can be used.  
• Only ingress points will be affected, but not the core of the network. |
| **User acceptance**       | • Users bid the true valuation (Vickrey), never pay more than they bid  
• Low-end users might get (possibly subsidised) cheap off-peak access | • Users purchase corresponding capacity if they want predictable capacity. No planning risk.  
• Low-end users can choose small "token bucket", slow rate and low frequency | • Higher quality of service is more expensive (but reduction through minimised packet size). Capacity is presold to minimise risk.  
• Lowest channel could be free |
| **Guarantee of service**  | • Relative priority established                                               | • No guarantee of service                                   | • Some guarantee of service through the pricing structure |
| **Capacity and congestion control** | • If the total value of packets exceeds costs of expanding the network, then capacity expansion will be feasible  
• Higher bids get access first, regardless of congestion | • Expansion requirements can be determined through accumulating statistics on the type of dropped OUT packets.  
• Packets are affected by the tagging mechanism only during congestion | • QoS based reservation of capacity can be accommodated within the lowest priced channel.  
• Self regulating congestion control based on feedback form user behaviour |

Hoe does flat rate pricing compare with pricing schemes which consider usage and quality of service? All proponents of differential pricing argue that flat rate pricing may have perverse consequences as maintaining sufficiently low price in a competitive
environment can result in revenues that are inadequate to recover costs (Varian, 1996; Sarkar, 1995). Sarkar argues strongly that the current practice of charging a flat rate fee for Internet access is “likely to severely impair” the very nature of the Internet. According to Sarkar, the packet-switched nature of the Internet is extremely vulnerable to congestion, and the cost of congestion has become a “tangible problem”.

Flat-rate pricing for Internet access seems to fail the main objectives of a pricing mechanism. First of all, it does not provide for an optimal allocation of the scarce resource of capacity. An example is provided by “bandwidth hungry” applications which need capacity reservation protocols; this in turn necessitate costs in developing edge functionality and at the same time decreases the overall efficiency of the network. And secondly, the flat rate pricing mechanism does not provide a feedback signal for future investment and capacity expansion.

On the other side, flat rate pricing has definite advantages – among them the simplicity of the accounting process, reduced costs for service providers, reduced transaction costs for users, a predictability of money flow (Odlyzko, 1997). A justification of the flat rate approach is offered in Anania & Solomon (1997), who argue that in an aggregated network the cost of switching is negligible. The differentiation between local, short-haul and long-distance network access provision disappears - in their words, “the carriage has merged with the commodity being carried”. That is why users need not pay for usage but only for access to the ingress point of their local ISP. They suggest that that dynamic allocation of network resources will become “increasingly difficult to meter and expensive to track by the carrier” Artificial costing separation will not reduce network congestion, and according to them the only feasible solution remains universal flat rate of access – paying in advance subscription being the best method of pricing in an integrated global digital network.

According to Odlyzko, flat rate pricing is a form of bundling and produces revenues that are significantly higher than volume charging. Even the PMP model – the simplest usage based pricing scheme proposed so far, does not have significant advantages
compared to variations of flat-rate pricing such as block pricing (providing the user with an allotment of data bytes), or “expected usage pricing” - unlimited access for a limited but sufficiently long period, taking into consideration the user’s profile (Odlyzko, 2001).

Another aspect of flat rate pricing is considered by Mason (2001) who points out that a pricing scheme, which is needed as tool to control Internet congestion, needs to balance both simplicity and robustness in a competitive environment. He shows that even a small fixed cost of implementing usage-based pricing can result in as situation where the best strategy for all competitors is flat rate pricing. Mason also shows that there might be no equilibrium: the market will cycle between flat rate and usage-based pricing.

What is the practical reality of Internet access pricing in New Zealand? As early as 1997, Brownlee (1997) describes usage based pricing in the New Zealand ‘s Internet gateway maintained by Waikato University. To study the current trends in the New Zealand Internet access provision industry, we consider the data on ISP pricing plans for July 2002 and February 2003 (Consumer Online, 2002; 2003).

Table 3. Internet access pricing in New Zealand

<table>
<thead>
<tr>
<th></th>
<th>July 2002</th>
<th>February 2003</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-tier: time based price</td>
<td>23 (14.56%)</td>
<td>26 (15.76%)</td>
<td>Some plans offers capped subscription</td>
</tr>
<tr>
<td>2-tier: the subscription covers a limited number of hours; when exceeded, a time based price is used</td>
<td>83 (52.53%)</td>
<td>96 (58.18%)</td>
<td>One plan differentiates the price in connection with regard to the time of the day (2002, 2003)</td>
</tr>
<tr>
<td>Flat rate</td>
<td>52 (32.91%)</td>
<td>36 (21.82%)</td>
<td>In 2003, two plans included a different price for more than 8 hours of continuous use</td>
</tr>
</tbody>
</table>

The overall number of ISPs in New Zealand is around 120, according to Bartley (2002). The number of large ISPs is significantly less – Telecom (Xtra) and TelstraClear (ClearNer and Paradise), which effectively makes New Zealand a duopoly in the realm of large ISPs. According to Consumer Online, in February 2003, thirty eight companies offered Internet access to “heavy” users, with a total number of 165 pricing plans (compared to 2002, when the same number of companies offered 156 different pricing plans). A “heavy user”
typically accesses the Internet for 150 hours each month (monthly average of 20 MB New Zealand traffic and 300 MB international traffic). The plans can be classified as 1-tier, 2-tier and flat rate, as shown in Table 3. During the relatively short period of six months, the number of plans offering usage-based pricing increased significantly, and consequently the number of flat rate plans went down. Odlyzko (2001) pointed out that whether to offer flat or metered rates in a competitive environment is a business decision. Flat rates encourage usage while usage-sensitive does not, but in the absence of conclusive evidence that the Internet is indeed endangered by congestion, he predicted that flat rate would remain the preferred pricing scheme in the nearest future. In New Zealand, the trend is just the opposite.

References


