Paris Metro Pricing: The minimalist differentiated services solution

Andrew Odlyzko
AT&T Labs - Research

***ABSTRACT*— Differentiated services for the Internet are undergoing intensive development. It is widely accepted that they will require usage sensitive pricing. The Paris Metro Pricing (PMP) proposal is to rely on pricing alone to provide differentiated services. PMP is the simplest differentiated services system in terms of complexity.**

1. Introduction

The Internet currently provides only best-effort service that treats all packets equally. However, there is wide dissatisfaction with the perceived performance, and there appears to be a wide consensus that new applications, especially real time ones such as packet telephony, will require changing how the Internet operates. Various QoS (quality of service) techniques are being developed and are beginning to be deployed. (For a general survey and references, see [4].) They will provide differentiated service levels. Many of these schemes are complicated, and involve substantial costs in both development and operations. Furthermore, since the basic problem is that of allocating a limited resource, it is widely accepted that all solutions will have to involve pricing mechanisms, to prevent users from sending all their traffic in the highest priority class.

I propose to simplify the problem by using simple pricing to provide congestion control. The proposal, called Paris Metro Pricing, or PMP, is to partition the main network into several logically separate channels. In the basic design, each would have a fixed fraction of the capacity of the entire network. (Many variations on this proposal are possible and some are discussed briefly in Section 2.) All channels would route packets using protocols similar to the current ones, with each packet treated equally. The only difference between the channels would be that they would charge different prices. Customers would choose the channel to send their packets on (possibly on a packet-by-packet basis), and would pay accordingly. There would be no formal guarantees of quality of service, with packets handled on a "best effort" basis. The expectation is that the channels with higher prices would be less congested than those with lower prices.

The PMP proposal was inspired by the Paris Metro system. Until about 15 years ago, when the rules were modified, the Paris Metro operated in a simple fashion, with 1st and 2nd class cars that were identical in number and quality of seats. The only difference was that 1st class tickets cost twice as much as 2nd class ones. (The Paris regional RER lines still operate on this basis.) The result was that 1st class cars were less congested, since only people who cared about being able to get a seat, etc., paid for 1st class. The system was self-regulating, in that whenever 1st class cars became too popular, some people decided they were not worth the extra cost, and traveled 2nd class, reducing congestion in 1st class and restoring the differential in quality of service between 1st and 2nd class cars.

Pricing is a crude tool. Different applications vary in requirements for bandwidth, latency, and jitter, for example. PMP would not provide any specific QoS guarantees. The justification for PMP is that, for all its deficiencies, the Internet does work, and with low congestion, even real-time applications run well. The main motivation behind PMP is to keep the Internet as simple as possible for the user.

The pricing mechanism of PMP is about as simple as that of any usage sensitive pricing scheme that has been proposed for the Internet. Thus the additional complexity it would introduce is minimal, and appears inevitable, since usage sensitive pricing appears inevitable. The advantage of PMP is that it would provide congestion control essentially for free, once the pricing mechanism is in place, with only minor changes to the network infrastucture being required to handle the traffic management tasks.

PMP is also designed to be acceptable to users, who have a strong preference for flat-rate pricing. It appears that consumers are willing to tolerate substantial variation in quality of a service or a product, but strongly prefer simple and predictable pricing schemes.

At a high level, PMP is similar to diff-serv, perhaps the most popular of the QoS techniques being developed. The difference is that diff-serv does not by itself say anything about assignment of priorities and pricing. It treats only the technical aspect of how the network should deal with packets with different markings. PMP integrates pricing with traffic management.

This position paper only outlines PMP. More details and references are available in the initial proposal [10]. For other references on pricing proposals, see the Web pages [3, 7, 14].

2. PMP

The main idea of PMP is simply to have several channels that differ in price. They would offer different expected quality of service through the action of users who select the channel to send their data on.

The number of channels in PMP should be small, possibly just two, but more likely three or four. Having few channels minimizes losses from not aggregating all the traffic, and also fits consumer preferences for simple schemes.

The basic version of PMP mentioned in the Introduction assigns to each channel a fixed fraction of the capacity of the entire network. One can also use weighted priorities, as in the weighted round-robin technique [4]. The advantage of the priority approach is that the full gain from aggregating all traffic on one network would be obtained. However, allowing high priority packets to block completely lower priority ones violates the fairness criterion that appears to be important to consumers.

In general, assignments of capacities and prices to the channels in PMP should stay constant for extended periods. This would fit consumer preferences for simplicity and also allow usage patterns to stabilize, and thus produce a predictable level of service on different channels. However, it would likely be desirable to have different assignments of capacities and prices for nights and weekends.

3. PMP problems and solutions

Would users find the lack of guaranteed quality of service (QoS) of PMP acceptable? In voice telephony, experience has taught people to expect a uniform and high level of service. However, that is an exception. Most purchases (of books, cars, and so on) are made on the basis of expected, not guaranteed, quality. Experimental networks such as vBNS, which have low utilization levels, are able to handle all applications. This suggests that PMP, a best-effort system without guarantees, but with several channels of different congestion levels, might satisfy most needs.

Would PMP survive in a competitive market? There is an analysis of a simplified version of PMP by Gibbens, Mason, and Steinberg [6] which shows that in their model, PMP would be optimal for a monopolist, but a carrier offering PMP would lose to one offering undifferentiated service. However, whether this analysis poses serious problems for PMP is not settled, since competition in information goods in general is hard to model, and most analyses predict destructive price wars (see [5], for example). Also, it has not been shown yet how any other QoS scheme can be realistically implemented in the Internet, which consists of many heterogeneous subnetworks.

The remainder of this section concentrates on a few aspects of PMP. The ability to assign varying capacities to the separate channels, and also to vary prices for using those channels gives service providers substantially more flexibility than might appear at first.

For the PMP proposal to work, the performance of the different channels has to be predictable, at least on average. Unfortunately, the fractal nature of data traffic means that we have to expect that all PMP channels will experience sporadic congestion. This could lead to network instability, with degradation on one channel propagating to other channels. There are several ways to overcome this problem (should it turn out to be a serious one). One is by modifying the charging mechanism. Access to the premium channels might be not on a packet-by-packet basis, but instead the user would pay for the right to send 1,000 packets on that channel in the next second. This would increase the financial barrier to upgrading channels. Block pricing could even be used in PMP to segregate streaming traffic from bursty data transfers.

4. PMP implementation

The PMP proposal can be regarded as a logical development of some current trends. A class of "premium ISPs" is developing, which provide higher quality of service. Customers with connections to several ISPs would then have a choice similar to that in PMP. The PMP proposal would simply let each ISP offer its customers an array of choices that they might have available through different ISPs anyway, and should therefore be more efficient.

PMP would be easy to introduce. It would not be necessary to wait for the deployment of IPv6 or other protocols. The current IPv4 packets already have a 3-bit priority field that is unused. Since the number of channels in PMP is likely not to exceed 4, this is more than sufficient. Interoperability would be easy, as all packets that do not contain any bits indicating class of service could be sent on the lowest cost (and lowest priority) channel.

At least initially, the cost per packet on the lowest cost channel would undoubtedly be zero. That would make this channel look like the current Internet, and so make the transition easier. It might also be possible to have zero prices on this channel in the long run during slack periods.

Inside the network, changes would only have to be done in the router software. It would be necessary to maintain logically separate queues or to give appropriate priority to packets from different channels. The current diff-serv QoS efforts in the IETF provide all the technical tools for implementing PMP.

The major change required in a network by PMP is the same one as that needed for any usage sensitive pricing scheme. It would be necessary to install hardware or software to count the packets and bytes for each user. Essentially all of this accounting could be done at the edges of the network.

Flat rates are preferred by consumers, but they also have major advantages for service providers. They were already advocated for broadband services by Anania and Solomon in [1], a paper that was first presented almost a decade ago. On the Internet, they eliminate the need for a traffic measurement and charging infrastructure, which, even for a system such as PMP, where almost all the work would be done at the edges of the network, would be costly to implement.

In PMP, the preference for flat-rate pricing can be partially accommodated by selling large blocks of transmission capacity (giving the user the right to send or receive 100 MB of data over a week through the lowest priced channel, or 60 MB through the next most expensive channel, say). Such pricing has worked well in long distance telephony in the United States, with consumers typically paying for more capacity than they used [9].

PMP offers a simple pricing plan with constant and easily understood pricing, which is an advantage, as it fits consumer desires. It does not offer any service guarantees, however. Such guarantees are popular. However, few guarantees are absolute, and most purchases are made on the basis of expectations. It seems likely that consumers could accept the lack of guarantees of QoS in PMP, especially if the average quality of different channels were predictable enough.

5. Applications of PMP

There are experts in the data networking community who argue that instead of working on complicated network schemes, all resources should be devoted to improving capacity (the "fat dumb pipe" model). The general consensus seems to be that this is not feasible, and that differentiated services are required to overcome the problem of "the tragedy of the commons," with rapid growth in traffic demand leading to endemic congestion. When I first proposed PMP [10], I shared this view, but based on knowledge of how many networks are operated, felt that one should strive for maximal simplicity even at the expense of maximal efficiency in use of transport capacity. A recent series of studies [2, 11, 12, 13] has raised questions about the basic assumptions that underlie the work on QoS, or at least the backbones of the Internet. Most of the Internet is very lightly utilized, most of the problems are not caused by link or switch congestion (which is what QoS measures address), and "the tragedy of the commons" is much less of a problem than is commonly believed. It appears that in the backbones of the Internet, providing a uniformly high quality of service to all transmissions might be not just feasible, but optimal, given the full cost that any QoS measures, even PMP, would impose. However, it is impossible to be certain this will be the case, since it is not clear how rapidly advances in transmission technology will translate into lower prices. If prices do not decline (and they have been rising in recent years), some QoS measures might be required even in the backbones. In that case, though, the studies mentioned above argue that nothing more complicated than PMP should be implemented. The reason is that networking is already too complicated. The behavior that has been observed (such as many network managers knowing practically nothing about the traffic on their networks, traffic staying on established private line networks instead of much less expensive Frame Relay services, and so on) shows that network managers already have too much to do, and it is unrealistic for them to assign proper priorities to different transmissions, say. Thus the arguments for maximal simplicity are very strong, and favor the use of PMP among all the differentiated service schemes if any QoS measures are required.

The arguments that QoS is unlikely to be needed because of rapid advances in photonics apply only to connections where fiber connections are feasible. There are bound to be many Internet connections (especially wireless ones, but also in local access through cable modem and ADSL) where available bandwidth will likely continue to be much more limited. In those contexts some QoS measures are likely to be required, and PMP may be helpful. For example, in Internet access through cable TV, instead of assigning a fixed number of households to each channel, one could have two channels with different prices, and let the households sort themselves out among them.

References

[1] L. Anania and R. J. Solomon, Flat–the minimalist price, pp. 91-118 in *Internet Economics,* L. W. McKnight and J. P. Bailey, eds., MIT Press, 1997. Preliminary version in *J. Electronic Publishing,* special issue on Internet economics, ⟨http://www.press.umich.edu/jep/⟩.

[2] K. G. Coffman and A. M. Odlyzko, The size and growth rate of the Internet, *First Monday,* 3(10) (October 1998), ⟨http://www.firstmonday.dk/⟩. Also available at ⟨http://www.research.att.com/∼amo⟩.

[3] N. Economides, The economics of networks, Web site with links to online sources, ⟨http://raven.stern.nyu.edu/networks/⟩.

[4] P. Ferguson and G. Huston, *Quality of Service: Delivering QoS on the Internet and in Corporate Networks,* Wiley, 1998.

[5] P. C. Fishburn, A. M. Odlyzko, and R. C. Siders, Fixed fee versus unit pricing for information goods: competition, equilibria, and price wars, *First Monday,* vol. 2, no. 7 (July 1997), ⟨http://www.firstmonday.dk/⟩. Available at ⟨http://www.research.att.com/∼amo⟩.

[6] R. Gibbens, R. Mason, and R. Steinberg, Multiproduct competition between congestible networks, available at ⟨http://www.soton.ac.uk/ ram2/papers.html⟩.

[7] J. MacKie-Mason, Telecom Information Resources on the Internet, Web site with links to online sources, ⟨http://china.si.umich.edu/telecom/telecom-info.html⟩.

[8] L. W. McKnight and J. P. Bailey, eds., *Internet Economics,* MIT Press, 1997. Preliminary version in *J. Electronic Publishing,* special issue on Internet economics, ⟨http://www.press.umich.edu/jep/⟩.

[9] B. M. Mitchell and I. Vogelsang, *Telecommunications Pricing: Theory and Practice,* Cambridge Univ. Press, 1991.

[10] A. M. Odlyzko, A modest proposal for preventing Internet congestion. Unpublished manuscript, available at ⟨http://www.research.att.com/∼amo⟩.

[11] A. M. Odlyzko, Data networks are lightly utilized, and will stay that way. Available at ⟨http://www.research.att.com/∼amo⟩.

[12] A. M. Odlyzko, The economics of the Internet: Utility, utilization, pricing, and Quality of Service. Available at ⟨http://www.research.att.com/∼amo⟩.

[13] A. M. Odlyzko, The Internet and other networks: Utilization rates and their implications. Available at ⟨http://www.research.att.com/∼amo⟩.

[14] H. R. Varian, The economics of the Internet, information goods, intellectual property and related issues, reference Web pages with links, ⟨http://www.sims.berkeley.edu/resources/infoecon/⟩.