Congestion Pricing

Paying Your Way in Communication Networks

Tristan Henderson, Jon Crowcroft, and Saleem Bhatti
University College London

Network congestion is a fundamental problem facing Internet users today. Many common complaints — "Why is this streaming video clip so jerky?" "Why are these Web pages taking so long to download compared to yesterday?" and "Why are my pings so high in Quake?" — usually reflect congestion somewhere in the network.

Traditional approaches to congestion control have viewed the Internet as a cooperative network. Transport protocols such as TCP were designed such that sources would adapt to network congestion by backing off and thus allow all sources to continue sending through the network, but at a reduced rate. Such congestion control is successful, however, only if users agree to cooperate in the manner mandated by the protocol designers. This has worked so far, perhaps because implementing a new transport protocol entails kernel programming that is slightly harder than virus-writing or other antisocial activities. In addition, some Internet service providers (ISPs) have forced their user traffic to conform to TCP-like behavior, such as that described in RFC 2581.[1]

On the other hand, denial-of-service attacks and demands for higher quality of service (QoS) are just two cases where network users prefer to act in their own self-interest rather than cooperate with others. A network where users are selfish, and thus reluctant to defer to other users, may result in the famous "tragedy of the commons," where — in the absence of controls — a shared resource is overconsumed by individuals who consider only their personal costs and not the cost to society as a whole.[2] In terms of the Internet, the "tragedy" could be viewed as congestive collapse, resulting from overconsumption of the shared network resource.

Economic Externalities

Economists define an *externality* as a cost (or benefit) of a good that does not accrue to the consumer of the good. Negative externalities became an economic issue in the early 20th century, when Arthur Pigou proposed a tax representing the difference between the marginal private cost and the marginal social cost of a good that exhibited externalities.[3] The Pigouvian tax would internalize externalities and force users to evaluate their costs to society before acting.

Figure 1 illustrates the Pigouvian tax. In the absence of intervention, users consider only their marginal private cost, MC_p, that is, the cost to the individual user of consuming an additional unit of the good. This leads to the consumption of quantity Q_1 at price P_1, as opposed to the socially optimal outcome of Q_2 at P_2. By imposing a tax of $P_3 - P_2$, individuals consider a new cost function MC_t, which results in the desired outcome, Q_2.

Externalities figure in many areas such as environmental pollution, deforestation, and — perhaps most relevant here — automobile traffic congestion. When drivers consider only the private costs of their automobile journeys, the use of popular roads at peak hours can trigger traffic jams. Many city governments have considered pricing schemes such as tolls to internalize these externalities.[4]

 1089-7801/01/$10.00

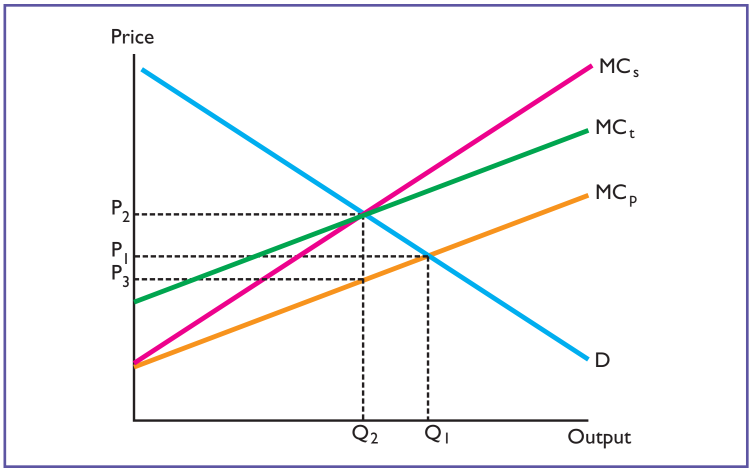

Figure 1. Internalizing externalities. The Pigouvian tax, $P_3 - P_2$ increases the cost observed by individuals from MC_p to MC_s to internalize the cost to society.

Network Pricing

It is important to distinguish congestion pricing from other forms of network pricing.

Most Internet users already pay for access. In the United States, users typically pay a monthly subscription to their ISP — a flat-rate charge that gives them little incentive to react to congestion or to consider the costs of their actions. In contrast, many other countries, particularly in Europe, feature usage-based pricing, where the price is related to the duration of the network connection. Such pricing plans can encourage users to disconnect from the Internet when they are not using it.

In the future, users may also have the option of paying for different levels of service. In the Diffserv model, ISPs can offer customers a range of QoS classes. This lets them differentiate prices for users who are willing to pay extra for a higher QoS, even though performance in an unloaded network may differ little between the service levels. Such price discrimination maximizes the provider's profits and is commonplace in other industries, most notably airplane travel, where adjacent travelers have seldom paid the same price for their seats.

Congestion Pricing

So where does congestion pricing fit into this framework? We know that charging network users for the congestion they cause can lead to more efficient network utilization by forcing them to take social costs into account. Yet this can seem counter-intuitive at first: Why should I be charged a congestion fee when I am actually receiving worse performance from the network and my ISP? In a congestion-pricing framework, however, the congestion charge would replace usage and QoS charges. Users would pay their ISPs a subscription charge to cover fixed costs, such as personnel and equipment, and a congestion charge only when appropriate. This pricing scheme is feasible because, in the absence of congestion, the marginal cost of a network link is practically zero. Once the link is built, additional traffic costs little.

Congestion pricing can also benefit network operators. By indicating the level of congestion and the user tolerance of it in their networks, congestion pricing can inform operators about when to reprovision and increase network capacity.

Smart Markets

Appropriately, economists were the first researchers to apply externalities and congestion pricing to the Internet. Hal Varian and Jeffrey Mackie-Mason proposed a "smart market" approach to allocating resources in a congested network.[5] At each congested router, a bid value in each ingress packet would indicate the amount that the packet's owner was willing to pay for the packet to pass through the router. The router would then hold an auction and admit the packets with the highest bids.

If the auctions are designed appropriately, the smart market can become an effective method of internalizing congestion externalities. For example, in a Vickrey auction, the winner pays not the highest, but the second-highest bid; this scheme is known to give users incentives to reveal their true preferences, since bidders are no longer scared of making too high a bid.[6]

If the auctions are designed appropriately, the smart market can internalize congestion externalities effectively. Unfortunately, this mechanism has been determined to be unworkable in a large internetwork. When there are multiple congested routers, a winning packet at the first congested router might lose at the second, because total bid value would now be decremented by the amount spent at the first router. To keep from losing money already spent on a congested route, a user might wish to program a packet with a new bid value at each router. But this would require network signaling in the reverse path to tell users the number of congested routers that had been traversed by each packet and the cost of each auction. This would not only increase traffic in an already-congested network but also further delay the very packets whose owners would consider paying a premium to traverse congested routers.

Edge Mechanisms

If pricing mechanisms at congestion points within the network do not scale, can a price calculation at the network's edge approximate an efficient solution? Network operators might prefer this, since they would retain control over how they charge users, rather than leaving it up to a network-mandated mechanism.[7] Instead of charging for the actual congestion caused by packets, the operators could charge for expected congestion, based on such metrics as the time of day, short-term congestion history, and so on.

The split-edge pricing framework[8] attempts to solve the problem of settling payments between interconnected domains. Instead of users making an individual payment to the owner of each congested router, split-edge pricing charges users only at the network's edge. Each network provider determines the cost of traffic traversing its network, and offers several classes of service to its neighbors at a set price. Both senders and receivers pay a charge for each transmission, and a clearinghouse iteratively settles the interdomain charges between them.

Statistical Approaches

Another approach is to make the routers aware of economic incentives, but not overload them with complex pricing mechanisms. Instead of dropping packets in the event of congestion, routers can mark them. The mark gives users an explicit signal of congestion, rather than the implicit signal they must infer from packet loss (which can be also be caused by transmission errors). Two bits in the IP header have been proposed for this mark, and are known as the explicit congestion notification (ECN) mechanism.

A network could use the ECN mark for congestion pricing as well. Since the mark indicates network congestion, the network can aggregate marks to represent a "shadow price" for the flow, reflecting the cost of the congestion it causes. Receivers can observe the number of marks, determine the charge for the flow, and then choose to continue or terminate the flow depending on this information.

By changing the marking or queuing algorithms used by routers, operators can optimize networks so that users acting selfishly to maximize their individual benefits will still contribute to a globally optimal resource allocation. A growing body of work examines potential marking algorithms and optimal allocation schemes to support goals such as "proportional fairness"[9] (flows should receive a share of the bandwidth proportional to the number of congested links they traverse) or "fair share"[10] (a router iteratively allocates an equal share of each class of service until each flow's bandwidth requirements are satisfied).

The Human Factor

All congestion-pricing schemes share one common element: the end users. It remains to be seen whether users will respond favorably to congestion pricing. The issue has been widely debated. In spite of its potential inefficiency, many users appear to favor flat-rate pricing. For example, judging from the American telephone market,[11] many residential users prefer a monthly inclusive fee for local calls that lets them know how much their bills will be in advance, even if they end up paying more than they would under a usage-based pricing scheme. When ISPs switch to a flat-rate scheme, network usage often increases.

Of course, the telephone network offers only a limited range of services, and users might be willing to pay premiums for particular network applications. Meanwhile, as the pricing debate runs on and on, the implementations of congestion pricing have thus far been few, and the user trials even fewer.

One project studied responses to bandwidth pricing by charging Internet users at the University of California, Berkeley, different amounts for access to different levels of bandwidth over ISDN connections.[12] Results from the Internet Demand Experiment (Index) project supported the idea that users might prefer flat-rate pricing, although the limited project scope did not include congestion- or QoS-based pricing.

The proportionally fair optimization solutions depend on logarithmic utility functions: as the amount of available bandwidth increases, the value that a user receives from the network increases logarithmically. Studies of user responses to variable network conditions indicate, however, that users prefer stable QoS levels.[13] This implies that users might prefer a fixed-bandwidth flow, even if a variable amount provided more bandwidth on average.

Some research suggests that the value users place on network performance (and so their willingness to deploy different utility-maximization strategies) may depend on the task at hand.[14] In a file transfer, a user might be interested only in how long it takes to deliver the last byte (assuming that the file is useless until this occurs). In this scenario, the session price for the entire file might be more important than the shadow prices of individual packets. On the other hand, in a streaming multimedia application, a user might want to maximize bandwidth share or minimize delay. Of course, as network line speeds increase, the bandwidth and delay goals become more difficult to fulfill concurrently and may require further development of

URLS OF INTEREST

- Band-X • http://www.band-x.com
- The Economics of Networks • http://www.stern.nyu.edu/networks/site.html
- Sally Floyd's page on end-to-end congestion control • http://www.aciri.org/floyd/tcpanalysis. html
- The M3I project • http://www.m3i.org
- Slashdot • http://slashdot.org
- The TCP-friendly Web site • http://www.psc.edu/networking/tcpfriendly.html
- Hal Varian's page on "The Information Economy" • http://www.sims.berkeley.edu/resources/infoecon/

router scheduler algorithms.

The cumulus pricing scheme[15] represents a compromise between flat-rate and usage-based pricing. Users pay a subscription fee that includes a base usage amount. Users who exceed the base amount receive red "cumulus points," and those who use less receive green points. At the end of each billing period, users pay extra for an account in the red or receive a refund for an account in the green. In an implementation, the cumulus points could perhaps reflect congestion pricing.

Implementation

Even if we assume that users will appreciate the benefits and accept congestion pricing, implementing a congestion pricing scheme still requires much work.

To create the appropriate incentives, congestion pricing requires a competitive Internet market, such that users can easily select alternatives, for example by changing routes or ISPs. Otherwise, a malicious ISP could deliberately generate congestion on a monopolistically-provided link, forcing its users to pay a congestion charge and thereby increasing the ISP's revenues. If users can avoid a congested link, they can defeat such a strategy. Although multihoming, where networks can choose to route between two or more upstream providers, has become a popular strategy for commercial networks, most end users currently have little control over the route their packets take. This situation is arguably becoming worse with the disappearance of the so-called "free" ISPs that charge only for telephone calls. Such ISPs facilitate user switching between access ISPs, since there are no contracts and little cost to changing ISP. Broadband connections such as cable modems and DSL lines, however, often entail long-term contracts that inhibit this switching.

In contrast, telephone markets appear to be giving users more options; for example, consider the countless long-distance connection choices and the contract-free, prepaid mobile phones that have become more popular than subscription phones in many countries.

Using ECN marks to indicate a shadow price creates its own implementation problems. ECN is in the process of becoming an IETF standard, not for pricing, but solely for the congestion notification that its label signifies. In particular, ECN is closely tied to a congestion-control router mechanism known as random early detection, or RED. Deployment of congestion pricing might prove difficult in an internetwork where various networks use ECN for congestion control, some use it for shadow pricing, and others ignore it altogether. Indeed, ECN has already experienced some deployment problems with buggy firewalls that reject ECN-capable flows. A firewall that alters the price of a flow through nonstandard responses to marks could have interesting economic consequences.

Many congestion pricing schemes assume a small number of long-term congested bottlenecks. While this might apply to relatively predictable road or telephone networks, Internet congestion is different. Most Internet congestion is unpredictable and can occur almost instantly at a popular site — such as the so-called "slashdot effect," named after the popular Web site that can generate huge increases in traffic to sites referenced in its news stories. Such congestion spikes could lead to highly unpredictable congestion prices.

What Next?

Though congestion pricing is a promising means of resource allocation in an evolving noncooperative network like the Internet, further research is required in many areas. Mobile networks, for example, present new pricing challenges. Neither multihop radio networks nor the increasingly popular, ad hoc wide-area 802.11b networks have any a priori infrastructure, which complicates accounting and resource allocation problems.

The slow deployment of IP multicast has been blamed on a lack of appropriate pricing and cost-recovery mechanisms for ISPs. Although this debate often fails to distinguish between pricing network usage and pricing the content to be delivered via multicast, there are several interesting issues raised by multicast congestion pricing: How should congestion charges be shared among the members of a multicast group? Can we design flexible mechanisms for apportioning charges between senders and receivers? How can we best fulfill user preferences if receivers have heterogeneous utility functions? Finally, multicast flows can potentially cause much more congestion than unicast traffic,[16] and this must be taken into account as well.

Perhaps the most important pricing consideration, however, is user perception. Many users seem reluctant to accept complex pricing mechanisms. Indeed, they may prefer a simpler, more predictable mechanism, even if it is not absolutely fair with respect to resource allocation. This may create a market for network capacity, where brokers absorb price fluctuations and offer users the simple view they prefer. The Market Managed Multiservice Internet (M3I) project is implementing a system for network resource management through market forces, and such brokers are one of the potential pricing solutions being examined. There are also signs of bandwidth being traded as a commodity at the network provider, as opposed to the user, level. For example, the electricity and communications supplier Enron is attempting to market wavelengths to corporate networks, and Band-X is a bandwidth exchange that lets companies tender out their network and bandwidth requirements to potential network providers. These developments indicate that even if congestion prices do not filter through to the average end-user's Internet usage bill, there might well be a place for congestion pricing elsewhere in the network.

References

1. M. Allman, V. Paxson, and W. R. Stevens, "TCP Congestion Control," Internet Engineering Task Force RFC 2581 (Informational), Apr. 1999; available online at http://www.ietf.org/rfcs/rfc2581.txt.
2. G. Hardin, "The Tragedy of the Commons," *Science*, vol. 162, Dec. 1968, pp. 1,243-1,248.
3. A.C. Pigou, *The Economics of Welfare*, Macmillan, London, 1920.
4. R. Arnott and K. Small, "The Economics of Traffic Congestion," *American Scientist*, vol. 82, no. 5, Sept.-Oct. 1994, pp. 446-455.
5. J.K. MacKie-Mason and H.R. Varian, "Pricing the Internet," *Public Access to the Internet*, B. Kahin and J. Keller, eds., MIT Press, Boston, Mass., 1995, pp. 269-314.
6. W. Vickrey, "Counterspeculation and Competitive Sealed Tenders", *J. Finance*, vol 16, no 1, March 1961, pp. 8-37.
7. S. Shenker et al., "Pricing in Computer Networks: Reshaping the Research Agenda," *Computer Comm. Rev.*, vol. 26, no. 2, Apr. 1996, pp. 19-43.
8. B. Briscoe, "The Direction of Value Flow in Connectionless Networks," *Proc. 1st Int'l Workshop on Networked Group Comm. (NGC)*, Springer-Verlag, Heidelberg, 1999, pp. 244-269.
9. F.P. Kelly, A.K. Maulloo, and D.K.H. Tan, "Rate Control in Communication Networks: Shadow Prices, Proportional Fairness and Stability," *J. Operational Research Soc.*, vol. 49, no. 3, Mar. 1998, pp. 237-252.
10. S.J. Shenker, "Making Greed Work in Networks: A Game-theoretic Analysis of Switch Service Disciplines," *IEEE/ACM Trans. on Networking*, vol. 3, no. 6, Dec. 1995, pp. 819-831.
11. A.M. Odlyzko, "Internet Pricing and the History of Communications," *Computer Networks*, vol. 36, no. 5-6, Aug. 2001, pp. 493-517.
12. R.J. Edell and P.P. Varaiya, "Providing Internet Access: What We Learn from the Index Trial," *IEEE Network*, vol. 13, no. 5, Sept./Oct. 1999, pp. 18-25.
13. A. Bouch, M.A. Sasse, and H. de Meer, "Of Packets and People: A User-centered Approach to Quality of Service," *Proc. 8th Int'l Workshop on Quality of Service (IWQoS)*, Springer-Verlag, Heidelberg, 2000, pp. 189-197.
14. P.B. Key and L. Massoulié, "User Policies in a Network Implementing Congestion Pricing," *Proc. Workshop on Internet Service Quality Economics (ISQE)*, MIT Press, Cambridge, Mass., 1999; available online at http://www.marengoresearch.com/isqe/.
15. P. Reichl et al, "How to Overcome the Feasibility Problem for Tariffing Internet Services: the Cumulus Pricing Scheme", *Proc. IEEE International Conference on Communications*, IEEE, Piscataway, NJ, 2001, pp 2079-2083.
16. A. Mankin et al., "IETF Criteria for Evaluating Reliable Multicast Transport and Application Protocols," IETF RFC 2357, June 1998; available online at http://www.ietf.org/rfc/rfc2357.txt.

Tristan Henderson is a PhD candidate in the Department of Computer Science, University College London. He has a BA (Hons) in economics from Emmanuel College, Cambridge University and an MSc in computer science from UCL. His research interests are network pricing and congestion control in multiuser network applications, and his doctoral work is examining quality of service considerations in multiplayer networked games.

Jon Crowcroft is a professor of networked systems in the Department of Computer Science, University College London. He has a bachelor's in physics from Trinity College, Cambridge University; and an MSc and PhD, both in computing, from UCL. He is a member of the ACM, a senior member of the IEEE, and a Fellow of the British Computer Society, the IEE, and the Royal Academy of Engineering. He is a member of the IAB and is on the editorial team for the *ACM/IEEE Trans. On Networks and Computer Communications*.

Saleem Bhatti is a lecturer in data communications and networking in the Dept. of Computer Science, University College London. He has a BEng (Hons) in electrical and electronic engineering, an MSc in data communication networks and distributed systems, and a PhD in computer science, all from UCL. His research interests are QoS (applications and networks), network management, network security, and mobile systems. He is a member of the IEEE and ACM.

Readers can contact the authors at {T.Henderson,J.Crowcroft,S.Bhatti}@cs.ucl.ac.uk.