Economic FAQs About the Internet

Jeffrey K. MacKie-Mason and Hal R. Varian

Abstract

This is a set of Frequently Asked Questions (and answers) about the economic, institutional, and technological structure of the Internet. We describe the history and current state of the Internet, discuss some of the pressing economic and regulatory problems, and speculate about future developments.

What is a FAQ?

FAQ stands for Frequently Asked Questions. There are dozens of FAQ documents on diverse topics available on the Internet, ranging from physics to scuba diving to how to contact the White House. They are produced and maintained by volunteers. This FAQ answers questions about the economics of the Internet (and towards the end offers some opinions and forecasts).

Where can the current version of this FAQ be found?

An earlier version of this FAQ was published in the Summer 1994 issue of the Journal of Economic Perspectives.[1] The Internet is changing at an astonishing rate so we have used this opportunity to revise and update the information in that earlier document. Future updates, can be found on the Web servers at the School of Public Policy at the University of Michigan [Telecom Information Directory] or the School of Information Management and Systems at UC Berkeley [The Information Economy].

BACKGROUND

What is the Internet?

The Internet is a world-wide network of computer networks that use a common communications protocol, TCP/IP (Transmission Control Protocol/Internet Protocol). TCP/IP provides a common language for interoperation between networks that use a variety of local protocols (Ethernet, Netware, AppleTalk, DECnet and others).

---

[1] We are grateful to the American Economics Association for permission to reprint substantial portions of that material.

Where did it come from?

In the late sixties, the Advanced Research Projects Administration (ARPA), a division of the U.S. Defense Department, developed the ARPAnet to link together universities and high-tech defense contractors. The TCP/IP technology was developed to provide a standard protocol for ARPAnet communications. In the mid-eighties the NSF created the NSFNET in order to provide connectivity to its supercomputer centers, and to provide other general services. The NSFNET adopted the TCP/IP protocol and provided a high-speed backbone for the developing Internet.

What do people do on the Internet?

Probably the most frequent use is electronic mail (e-mail). After that are file transfer (moving data from one computer to another) and remote login (logging into a computer that is running somewhere else on the Internet). In terms of traffic volume, as of December 1994 about 32% of total traffic was file transfer, 16% was World Wide Web (WWW), 11% was netnews, 6% was email, 4% was gopher, and the rest was for other uses [Merit Statistics]. People can search databases (including the catalogs of the Library of Congress and scores of university research libraries), download data and software, and ask (or answer) questions in discussion groups on numerous topics (including economics research).

How big is the Internet?

From 1985 to December 1994, the Internet grew from about 200 networks to well over 45,000 and from 1,000 hosts (end-user computers) to over four million. About 1,000,000 of these hosts are at educational sites, 1,300,000 are commercial sites, and about 385,000 are government/military sites, all in the U.S. Most of the other 1,300,000 hosts are elsewhere in the world [Network Wizards]. NSFNET traffic grew from 85 million packets in January 1988 to 86 billion packets in November 1994. (Packets are variable in length, with a bimodal distribution. The mean is about 200 bytes on average, and a byte corresponds to one ASCII character.) This is more than a six hundred-fold increase in only six years. The traffic on the network is currently increasing at a rate of 6% a month. (NSFNET statistics are available at Merit's Network Information Center[1].)

John Quarterman estimates that as of October 1994 there were about 8 million people directly connected to the Internet, and about 6 million more people who can access the Internet "indirectly'' through online services. The numbers in the latter category are likely substantially larger today. Of course the total number of people who have access to the Internet via email is

[1] Beginning in the early 1990s, the statistics do not reflect the size of the total U.S. network because alternative backbones began appearing. It is generally believed that the NSFNET accounted for at least 75% of U.S. backbone traffic until around September 1994, after which its share rapidly fells as the NSFNET was gradually phased out. Shutdown occurred on 30 April 1995.

much larger---it may even approach the 20--30 million figures bandied about in the mass media.

ORGANIZATION

Who runs the Internet?

The short answer is "no one.'' The Internet is a loose amalgamation of computer networks run by many different organizations in over seventy countries. Most of the technological decisions are made by small committees of volunteers who set standards for interoperability.

What is the structure of the Internet?

The US portion of the Internet is best thought of as having three levels. At the bottom are local area networks (LANs); for example, campus networks. Usually the local networks are connected to a regional, or mid-level network. The mid-levels connect to one or more backbones. A backbone is an overarching network to which multiple regional networks connect, and which generally does not serve directly any local networks or end-users. The U.S. backbones connect to other backbone networks around the world. There are, however, numerous exceptions to this structure.

A few years ago the primary backbone was the NSFNET. On April 30, 1995 the NSFNET ceased operation and now traffic in the US is carried on several privately operated backbones. The new "privatized Internet'' in the US is becoming less hierarchical and more interconnected. The separation between the backbone and regional network layers of the current structure are blurring, as more regionals are connected directly to each other through network access points (NAPs), and traffic passes through a chain of regionals without any backbone transport.

What are the backbone networks?

In January 1994 there were four public fiber-optic backbones in the U.S.: NSFNET, Alternet, PSInet, and SprintLink. The NSFNET was funded by the NSF; it evolved directly out of ARPANET, the original TCP/IP network. The other backbones were private, for-profit enterprises.

By summer 1995 there were at least 14 national and super-regional high-speed TPC/IP networks in the U.S. As interconnection proliferates, this distinction becomes less important. A map of the major interconnection points and the numerous networks that use them is available at [CERFnet].

MCI, which helped operate the original NSFNET, is probably the largest carrier of Internet

traffic today; it claims to carry 40% of all Internet traffic. However, this is a highly competitive market; Sprint, Alternet, and PSInet are also signing up many customers.

What was the NSFNET?

The NSFNET was the first backbone for the US portion of the Internet. It was originally conceived as a way for researchers to submit jobs to supercomputers located at various universities around the US. Subsequently it was realized that the excess capacity on this backbone could be used to exchange data among universities that had nothing to do with supercomputing. The NSF annually paid about $11.5 million for the NSFNET operation for several years, but eventually decided that the technology was mature enough that it could be more effectively provided by the private market.

What happened to the NSFNET?

The NSFNET backbone was shut down on April 30, 1995, as the NSF funding for it ended. NSF is continuing to fund some regional nets, but this funding steadily decreases to zero over five years. Instead, the NSF is funding Network Access Points (NAPS) near Chicago, San Francisco, and New York. The NAPs are interconnection points for backbone providers. See [Fazio 1995] for an article describing the transition in detail; current information is available at the Merit Web site for information on the transition [Merit Architecture]. The NSF is also funding a routing arbiter service to provide fair and efficient routing among the various backbones and regionals.

The NSF is also funding the vBNS (very-high speed backbone network service) to connect five of its supercomputer sites at 155 Mbps. Its emphasis will be on developing capabilities for high-definition remote visualization and video transmission.

How much did NSFNET cost?

It is difficult to say how much the Internet as a whole costs, since it consists of thousands of different networks, many of which are privately owned. However, it is possible to estimate the cost of the NSFNET backbone, since it was publicly supported. In 1993, NSF paid Merit about $11.5 million per year to run the backbone. Approximately 80% of this was spent on lease payments for the fiber optic lines and routers. About 7% of the budget was spent on the Network Operations Center, which monitor traffic flows and troubleshoots problems.

To give some sense of the scale of this subsidy, add to it the approximately $7 million per year that NSF paid to subsidize various regional networks, for a total of about $20 million. Based on estimates that there were approximately 20 million Internet users (most of whom were connected to the NSFNET in one way or another), the NSF subsidy amounted to about $1 per user per year.

Of course, this was significantly less than the total cost of the Internet; indeed, it does not even include all of the public funds, which came from state governments, state-supported universities, and other national governments as well. No one really knows how much all this adds up to, although there are some research projects underway to try to estimate the total U.S. expenditures on the Internet. It has been estimated---read "guessed''--- that the NSF subsidy of $20 million per year was less than 10% of the total by expenditure U.S. public agencies on the Internet.

Who provides access outside of the U.S.?

There are now a large number of backbone and mid-level networks in other countries. For example, most western European countries have national networks that are attached to EBone, the European backbone. The infrastructure is still immature, and quite inefficient in some places. For example, the connections between other countries often are slow or of low quality, so it was common to see traffic between two countries that is routed through the NSFNET in the U.S. [Braun and Claffy 1993].

TECHNOLOGY

Is the Internet different from telephone networks?

Yes and no. Most backbone and regional network traffic moves over leased phone lines, so at a low level the technology is the same. However, there is a fundamental distinction in how the lines are used by the Internet and the phone companies. The Internet provides connectionless packet-switched service whereas telephone service is circuit-switched. (We define these terms below.) The difference may sound arcane, but it has vastly important implications for pricing and the efficient use of network resources.

What is circuit-switching?

Phone networks use circuit switching: an end-to-end circuit must be set up before the call can begin. A fixed share of network resources is reserved for the call, and no other call can use those resources until the original connection is closed. This means that a long silence between two teenagers uses the same resources as an active negotiation between two fast-talking lawyers. One advantage of circuit-switching is that it enables performance guarantees such as guaranteed maximum delay, which is essential for real-time applications like voice conversations. It is also much easier to do detailed accounting for circuit-switched network usage.

How is packet-switching technology different from circuit-switching?

The Internet uses "packet-switching'' technology. The term "packets'' refers to the fact that the data stream from your computer is broken up into packets of about 200 bytes (on average),

which are then sent out onto the network.[1] Each packet contains a "header'' with information necessary for routing the packet from origination to destination. Thus each packet in a data stream is independent.

The main advantage of packet-switching is that it permits “statistical multiplexing” or “statistical sharing” on the communications lines. That is, the packets from many different sources can share a line, allowing for very efficient use of the fixed capacity. With current technology, packets are generally accepted onto the network on a first-come, first-served basis. If the network becomes overloaded, packets are delayed or discarded ("dropped'').

How are packets routed to their destination?

The Internet protocol is connectionless.[2] This means that there is no end-to-end setup for a session; each packet is independently routed to its destination. When a packet is ready, the host computer sends it on to another computer, known as a router. The router examines the destination address in the header and passes the packet along to another router, chosen by a route-finding algorithm. A packet may go through 30 or more routers in its travels from one host computer to another. Because routes are dynamically updated, it is possible for different packets from a single session to take different routes to the destination.

Along the way packets may be broken up into smaller packets, or reassembled into bigger ones. When the packets reach their final destination, they are reassembled at the host computer. The instructions for doing this reassembly are part of the TCP/IP protocol suite.

Some packet-switching networks are "connection-oriented'' (notably, X.25 networks, such as Tymnet and frame-relay networks). In such a network a connection is set up before transmission begins, just as in a circuit-switched network. A fixed route is defined, and information necessary to match packets to their session and defined route is stored in memory tables in the routers. Thus, connectionless networks economize on router memory and connection set-up time, while connection-oriented networks economize on routing calculations (which have to be repeated for every packet in a connectionless network).

What is the physical technology of the Internet?

Most of the network hardware in the Internet consists of communications lines and switches or routers. In the regional and backbone networks, the lines are mostly leased telephone trunk lines, which are increasingly fiber optic. Routers are computers; indeed, the routers used on the

[1] Recall that a byte is equivalent to one ASCII character.

[2] TCP is a connection-oriented protocol that is overlaid on the IP protocol by most, but not all applications.

NSFNET were modified commercial IBM RS6000 workstations, although custom-designed routers by other companies such as Cisco, Wellfleet, 3-Com and DEC probably have the majority share of the market.

What does "speed'' mean?

"Faster'' networks do not move electrons or photons at faster than the speed of light; a single bit travels at essentially the same speed in all networks. Rather, "faster'' refers to sending more bits of information simultaneously in a single data stream (usually over a single communications line), thus delivering n bits faster. Phone modem users are familiar with recent speed increases from 300 bps (bits per second) to 2400, 9600 and now 19,200 bps. Leased-line network speeds have advanced from 56 Kbps (kilo, or 10^3 bps) to 1.5 Mbps (mega, or 10^6 bps, known as T-1 lines) in the late 80s, and then to 45 Mbps (T-3) in the early 90s. Lines of 155 Mbps are now available, though not yet widely used. The U.S. Congress had called for a 1 Gbps (giga, or 10^9] bps) backbone by 1996. This goal has been nearly achieved in testbeds, though it now looks like it will be at least a couple of more years before we see gigabit speeds in the public backbone.

Current T-3 45 Mbps lines can move data at a speed of 1,400 pages of text per second; a 20-volume encyclopedia can be sent coast to coast in half a minute. However, it is important to remember that this is the speed on the wide area network---the "access roads" via the regional networks still mostly use the much slower T-1 connections.[1]

Why do data networks use packet-switching?

Economics can explain most of the preference for packet-switching over circuit-switching in the Internet and other public networks. Circuit networks use lots of lines in order to economize on switching and routing. That is, once a call is set up, a line is dedicated to its use regardless of its rate of data flow, and no further routing calculations are needed. This network design makes sense when lines are cheap relative to switches.

The costs of both communications lines and computers have been declining exponentially for decades. However, since about 1970, switches (computers) have become relatively cheaper than lines. At that point packet switching became economic: lines are shared by multiple connections at the cost of many more routing calculations by the switches. This preference for using many relatively cheap routers to manage few expensive lines is evident in the topology of the backbone networks. For example, in the NSFNET any packet coming on to the backbone had to pass through two routers at its entry point and again at its exit point. A packet entering at Cleveland and exiting at New York traversed four routers but only one leased T-3 communications line.

---

1 As with almost everything in this FAQ, the details are constantly changing. As of the final editing of this version (May 1996), some of the Internet backbone links are running at 155 Mbps.

What are ATM and cell-switching technologies?

The international telephone community has committed to a future network design that combines elements of both circuit and packet switching to enable the provision of integrated services. The ITU (an international standards body for telecommunications, formerly known as the CCITT) has adopted a "cell-switching'' technology called ATM (asynchronous transfer mode) for future high-speed networks. Cell switching closely resembles packet switching in that it breaks a data stream into packets which are then placed on lines that are shared by several streams. One major difference is that cells have a fixed size while packets can have different sizes. This makes it possible in principle to offer bounded delay guarantees (since a cell will not get stuck for a surprisingly long time behind an unusually large packet).

An ATM network also resembles a circuit-switched network in that it provides connection-oriented service. Each connection has set-up phase, during which a "virtual circuit'' is created. The fact that the circuit is virtual, not physical, provides two major advantages. First, it is not necessary to reserve network resources for a given connection; the economic efficiencies of statistical multiplexing can be realized. Second, once a virtual circuit path is established switching time is minimized, which allows for much higher network throughput. Initial ATM networks are already being operated at 155 Mbps, while the non-ATM Internet backbones operate at no more than 45 Mbps. The path to 1000 Mbps (gigabit) networks seems much clearer for ATM than for traditional packet switching.

What changes are likely in network technology?

At present there are many overlapping information networks (e.g., telephone, telegraph, data, cable TV), and new networks are emerging rapidly (paging, personal communications services, etc.). Each of the current information networks was engineered to provide a particular type of service and the added value provided by each different type was sufficient to overcome the fixed costs of building overlapping physical networks.

However, given the high fixed costs of providing a network, the economic incentive to develop an "integrated services'' network is strong. Furthermore, now that all information can be easily digitized separate networks for separate types of traffic are no longer necessary. Convergence toward a unified, integrated services network is a basic feature in most visions of the much publicized "information superhighway'' (e.g., [National Academy of Sciences 1994]). The migration to integrated services networks will have important implications for market structure and competition.

When will the "information superhighway'' arrive?

The federal High Performance Computing Act of 1991 aimed for a gigabit per second (Gbps) national backbone by 1995. Five federally-funded testbed networks are currently demonstrating various gigabit approaches. To get a feel for how fast a gigabit per second is, note that most small colleges or universities today have 56 Kbps Internet connections. At 56 Kbps it takes about five hours to transmit one gigabit!

Efforts to develop integrated services networks also have exploded. Several cable companies have already started offering Internet connections to their customers.[1] AT&T, MCI and all of the "Baby Bell'' operating companies are involved in mergers and joint ventures with cable TV and other specialized network providers to deliver new integrated services such as video-on-demand. ATM-based networks, although initially developed for phone systems, ironically have been first implemented for data networks within corporations and by some regional and backbone providers.

HOW IS INTERNET ACCESS PRICED?

What types of pricing schemes are used?

Until recently, nearly all users faced the same pricing structure for Internet usage. A fixed-bandwidth connection was charged an annual fee, which allowed for unlimited usage up to the physical maximum flow rate (bandwidth). We call this "connection pricing''. Most connection fees were paid by organizations (universities, government agencies, etc.) and the users paid nothing themselves.

Simple connection pricing still dominates the market, but a number of variants have emerged. The most notable is "committed information rate'' pricing. In this scheme, an organization is charged a two-part fee. One fee is based on the bandwidth of the connection, which is the maximum feasible flow rate; the second fee is based on the maximum guaranteed flow to the customer. The network provider installs sufficient capacity to simultaneously transport the committed rate for all of its customers, and installs flow regulators on each connection. When some customers operate below that rate, the excess network capacity is available on a first-come, first-served basis for the other customers. This type of pricing is more common in private networks than in the Internet because a TCP/IP flow rate can be guaranteed only network by network, greatly limiting its value unless a large number of the 20,000 Internet networks coordinate on offering this type of guarantee.

[1] Because most cable networks are one-way, many of these initial efforts use an "asymmetric'' network connector that brings the input in through the TV cable at 10 Mbps, but sends the output out through a regular phone line at about 14.4 Kbps. This scheme may be popular since most users tend to download more information than they upload.

Networks that offer committed information pricing generally have enough capacity to meet the entire guaranteed bandwidth. This is a bit like a bank holding 100% reserves in case all depositors want to withdraw on the same day. However, full provisioning is necessary with existing TCP/IP network technology since there is no commonly used way to prioritize packets, and because the statistical fluctuations in traffic are huge.

For most usage, the marginal packet placed on the Internet is priced at zero. At the outer fringes there are a few exceptions. For example, several private networks (such as Compuserve) provide email connections to the Internet. Several of these charge per message above a low threshold. The public networks in Chile [Baeza-Yates et al. 1993] and New Zealand [Brownlee 1994, 1996] charge their customers by the packet for all international traffic. An economic study of the New Zealand system can be found in [Carter and Guthrie 1994].

What other types of pricing have been considered?

Standard economic theory suggests that prices should be matched to costs. There are three main elements of network costs: the cost of connecting to the net, the cost of providing additional network capacity, and the social cost of congestion. Once capacity is in place, direct usage cost is negligible, and by itself is almost surely is not worth charging for given the accounting and billing costs (see [MacKie-Mason and Varian 1995b]).

Charging for connections is conceptually straightforward: a connection requires a line, a router, and some labor effort. The line and the router are reversible investments and thus are reasonably charged for on annual lease basis (though many organizations buy their own routers). Indeed, this is essentially the current scheme for Internet connection fees.

Charging for incremental capacity requires usage information. Ideally, we need a measure of the organization's demand during the expected peak period of usage over some period, to determine its share of the incremental capacity requirement. In practice, it might seem that a reasonable approximation would be to charge a premium price for usage during pre-determined peak periods (a positive price if the base usage price is zero), as is routinely done for electricity. However, casual evidence suggests that peak demand periods are much less predictable than for other utility services. One reason is that it is very easy to use the computer to schedule some activities for off-peak hours, leading to a shifting peaks problem.[1] In addition, so much traffic traverses long distances around the globe that time zone differences are important. Network statistics reveal very irregular time-of-day usage patterns [MacKie-Mason and Varian 1995a].

[1] The single largest current use of network capacity is file transfer, much of which is distribution of files from central archives to distributed local archives. The timing for a large fraction of file transfer is likely to be flexible. Just as most fax machines allow faxes to be transmitted at off-peak times, large data files could easily be transferred at off-peak times---if users had appropriate incentives to adopt such practices.

How can the Internet deal with increasing congestion?

If you have read this far in the chapter, you should have a good basic understanding of the current state of the Internet---we hope that most of the questions you have had about the how the Internet works have been answered. Starting here we will move from FAQs and "facts'' towards conjectures, FEOs (firmly expressed opinions), and PBIs (partially baked ideas). We first discuss congestion problems.

Nearly all usage of the Internet backbones is unpriced at the margin. Organizations pay a fixed fee in exchange for unlimited access up to the maximum throughput of their particular connection. This is a classic problem of the commons. The externality exists because a packet-switched network is a shared-media technology: each extra packet that Sue User sends imposes a cost on all other users because the resources Sue is using are not available to them. This cost can come in form of delay or lost (dropped) packets.

Without an incentive to economize on usage, congestion can become quite serious. Indeed, the problem is more serious for data networks than for many other congestible resources because of the tremendously wide range of usage rates. On a highway, for example, at a given moment a single user is more or less limited to putting either one or zero cars on the road. In a data network, however, single user at a modern workstation can send a few bytes of e-mail or put a load of hundreds of Mbps on the network. Today any undergraduate with a new Macintosh is able to plug in a digital video camera and transmit live videos to another campus or home to mom, demanding as much as 1 Mbps. Since the maximum throughput on current backbones is only 45 Mbps, it is clear that even a few users with relatively inexpensive equipment could bring the network to its knees.

Congestion problems are not just hypothetical. For example, congestion was quite severe in 1987 when the NSFNET backbone was running at much slower transmission speeds (56 Kbps) [Bohn et al. 1993]. Users running interactive remote terminal sessions were experiencing unacceptable delays. As a temporary fix, the NSFNET programmed the routers to give terminal sessions (using the telnet program) higher priority than file transfers (using the ftp program).

More recently, many services on the Internet have experienced severe congestion problems. Large ftp archives, Web servers at the National Center for Supercomputer Applications, the original Archie site at McGill University and many services have had serious problems with overuse. See [Markoff 1993] for more detailed descriptions.

Congestion on the trans-Atlantic link, which has been only 6 Mbps, has been quite severe, causing researchers requiring substantial bandwidth to schedule their work during the wee hours.

Since the advent of WWW and CU-SeeMe video-conferencing, there has also been seriously disruptive congestion in Europe. Indeed, for a period beginning in 1995, EUnet (the main European Internet backbone) forbade the use of CU-SeeMe without permission in advance.

If everyone just stuck to ASCII email congestion would not likely become a problem for many years, if ever. However, the demand for multi-media services is growing dramatically. Although the supply of bandwidth is increasing dramatically, so is the demand. If congestion remains unpriced it is likely that there will be increasingly damaging episodes when the demand for bandwidth exceeds the supply in the foreseeable future.

What non-price mechanisms can be used for congestion control?

Administratively assigning different priorities to different types of traffic is appealing, but impractical as a long-run solution to congestion costs due to the usual inefficiencies of rationing. However, there is an even more severe technological problem: it is impossible to enforce. From the network's perspective, bits are bits and there is no certain way to distinguish between different types of uses. By convention, most standard programs use a unique identifier that is included in the TCP header (called the "port'' number); this is what NSFNET used for its priority scheme in 1987. However, it is a trivial matter to put a different port number into the packet headers; for example to assign the telnet number to ftp packets to defeat the 1987 priority scheme. To avoid this problem, NSFNET kept its prioritization mechanism secret, but that is hardly a long-run solution.

What other mechanisms can be used to control congestion? The most obvious approach for economists is to charge some sort of congestion price. However, to date, there has been almost no serious consideration of congestion pricing for backbone services, and even tentative proposals for usage pricing have been met with strong opposition. We will discuss pricing below but first we examine some non-price mechanisms that have been proposed.

Many proposals rely on voluntary efforts to control congestion. Numerous participants in congestion discussions suggest that peer pressure and user ethics will be sufficient to control congestion costs. For example, recently a single user started broadcasting a 350--450 Kbps audio-video test pattern to hosts around the world, blocking the network's ability to handle a scheduled audio broadcast from a Finnish university. A leading network engineer sent a strongly-worded e-mail message to the user's site administrator, and the offending workstation was disconnected from the network. However, this example also illustrates the problem with relying on peer pressure: the inefficient use was not terminated until after it had caused serious disruption. Further, it apparently was caused by a novice user who did not understand the impact of what he had done; as network access becomes ubiquitous there will be an ever-increasing number of unsophisticated users who have access to applications that can cause severe congestion if not properly used. And of course, peer pressure may be quite ineffective against

malicious users who want to intentionally cause network congestion.

One recent proposal for voluntary control is closely related to the 1987 method used by the NSFNET [Bohn et al. 1993]. This proposal would require users to indicate the priority they want each of their sessions to receive, and for routers to be programmed to maintain multiple queues for each priority class. Obviously, the success of this scheme would depend on users' willingness to assign lower priorities to some of their traffic. In any case, as long as it is possible for just one or a few abusive users to create crippling congestion, voluntary priority schemes that are not robust to forgetfulness, ignorance, or malice may be largely ineffective.

In fact, a number of voluntary mechanisms are in place today. They are somewhat helpful in part because most users are unaware of them, or because they require some programming expertise to defeat. For example, most implementations of the TCP protocols use a "slow start'' algorithm which controls the rate of transmission based on the current state of delay in the network. Nothing prevents users from modifying their TCP implementation to send full throttle if they do not want to behave "nicely.''

A completely different approach to reducing congestion is purely technological: overprovisioning. Overprovisioning means maintaining sufficient network capacity to support the peak demands without noticeable service degradation.[1] This has been the most important mechanism used to date in the Internet. However, overprovisioning is costly, and with both very-high-bandwidth applications and near-universal access fast approaching, it may become too costly. In simple terms, will capacity demand grow faster than the decline in capacity cost?

Given the explosive growth in demand and the long lead time needed to introduce new network protocols, the Internet may face serious problems very soon if productivity increases do not keep up. Therefore, we believe it is time to seriously examine incentive-compatible allocation mechanisms, such as various forms of congestion pricing.

Can bandwidth be reserved?

The current Internet offers a single service quality: "best efforts packet service.'' Packets are transported first-come, first-served with no guarantee of success. Some packets may experience severe delays, while others may be dropped and never arrive.

However, different kinds of data place different demands on network services. E-mail and file transfers requires 100% accuracy, but can easily tolerate delay. Real-time voice broadcasts require much higher bandwidth than file transfers, and can only tolerate minor delays, but they can tolerate significant distortion. Real time video broadcasts have very low tolerance for delay

[1] The effects of network congestion are usually negligible until usage is very close to capacity.

and distortion.

Voice telephony networks handle the quality of service problem by assigning each call a physical circuit with fixed resources sufficient to guarantee a minimal quality of service. One limitation of this scheme is that the amount of resources devoted to each call is hardwired into the engineering of the network. As we have discussed, the Internet takes the approach of sharing all of its resources, all of the time, which accommodates the wildly varying bandwidth requirements of different applications, and which gains from averaging over the wildly varying bandwidth requirements during a session with most applications. A hybrid approach to offering different resources, along with guarantees, to different uses would be to allow flexible, dynamic resource reservation. That is, allow a user (or her software agent) to declare how much bandwidth, what maximal delay and what type of delay variation she requires for a given session, and allocate those resources to her. An experimental implementation of such a scheme is given in the RSVP protocol [Zhang et al. 1993].

How can users be induced to choose the right level of service?

Because of different resource requirements, network efficiency can be increased if the different types of traffic are treated differently---giving a delay guarantee to, say, a real-time video session but not to routine e-mail or file transfer. But in order to do this, the user must truthfully indicate what type of traffic he or she is sending. If real-time video bit streams get the highest quality service, why not claim that all of your bit streams are real-time video?

[Cocchi et al. 1992] point out that it is useful to look at network pricing as a mechanism design problem. The user can indicate the "type'' of his transmission, and the workstation in turn reports this type to the network. In order to ensure truthful revelation of preferences, the reporting and billing mechanism must be incentive compatible. The field of mechanism design has been criticized for ignoring bounded rationality of human subjects. However, in this context, the workstation is doing most of the computation, so that quite complex mechanisms may be feasible.

Why should pricing be taken seriously for congestion control?

Let us turn this question on its head: Why should data network usage be free even to universities, when telephone and postal usage are not?[1] The question is, does society benefit more from

[1] Many university employees routinely use email rather than the phone to communicate with friends and family at other Internet-connected sites. Likewise, a service is now being offered to transmit faxes between cities over the Internet for free, then paying only the local phone call charges to deliver them to the intended fax machine. And during early 1985, several versions of Internet voice telephone software have been released, allowing people to hold two-way conversations --- using large amounts of bandwidth --- but paying nothing to offset the service quality degradation they are imposing on other users.

priced or unpriced network resources?

As we have argued, other approaches to controlling congestion are either flawed or have undesirable side-effects. Pricing approaches have the overwhelming advantage that they permit users, acting individually (or as organizations, if the pricing is only applied at the organizational level) to express the value that they place on obtaining network services. Thus, pricing directly provides the information needed to allocate scarce resources during times of congestion to those users who value them most. There is no need to assign arbitrary priorities, or to force high-value users to suffer from being stuck in a first-come, first-served line behind low-value users. The paper by [MacKie-Mason et al. 1996a] in this volume, and [MacKie-Mason and Varian 1995c] discuss the advantages of pricing for congestion in more detail.

How might prices be used to control congestion?

We have elsewhere described a scheme for efficient pricing of the congestion costs ([MacKie-Mason and Varian 1995b], [MacKie-Mason and Varian 1995a]). The basic problem is that when the network is near capacity, a user's incremental packet imposes costs on other users in the form of delay or dropped packets. Our scheme for internalizing this cost is to impose a congestion price on usage that is determined by a real-time Vickrey auction. Following the terminology of Vernon Smith and Charles Plott, we call this a "smart market.''

The basic idea is simple. Much of the time the network is uncongested, and the price for usage should be zero. When the network is congested, packets are queued and delayed. The current queuing scheme is FIFO. We propose instead that packets should be prioritized based on the value that the user puts on getting the packet through quickly. To do this, each user assigns her packets a bid measuring her willingness-to-pay for immediate servicing. At congested routers, packets are prioritized based on bids. In order to make the scheme incentive-compatible, users are not charged the price they bid, but rather are charged the bid of the lowest priority packet that is admitted to the network. It is well-known that this mechanism provides the right incentives for truthful revelation.

This scheme has a number of nice features. In particular, not only do those with the highest cost of delay get served first, but the prices also send the right signals for capacity expansion in a competitive market for network services. If all of the congestion revenues are reinvested in new capacity, then capacity will be expanded to the point where its marginal value is equal to its marginal cost.[1]

[1] See [Gupta et al. 1994, 1996] for a related study of priority pricing to manage Internet congestion.

What are some problems with a smart market?

Prices in a real-world smart market cannot be updated continuously. The efficient price is determined by comparing a list of user bids to the available capacity and determining the cutoff price. In fact, packets arrive not all at once but over time, and thus it would be necessary to clear the market periodically based on a time-slice of bids. The efficiency of this scheme, then, depends on how costly it is to frequently clear the market and on how persistent the periods of congestion are. If congestion is exceedingly transient then by the time the market price is updated the state of congestion may have changed.[1]

A number of network specialists have suggested that many customers---particularly not-for-profit agencies and schools---will object because they do not know in advance how much network utilization will cost them. We believe that this argument is partially a red herring, since the user's bid always controls the maximum that network usage costs. Indeed, since we expect that for most traffic the congestion price will be zero, it should be possible for most users to avoid ever paying a usage charge by simply setting all packet bids to zero.[2] When the network is congested enough to have a positive congestion price, these users will pay the cost in units of delay rather than cash, as they do today.

We also expect that in a competitive market for network services, fluctuating congestion prices would usually be a "wholesale'' phenomenon, and that intermediaries would repackage the services and offer them at a guaranteed price to end-users. Essentially this would create a futures market for network services.

There are also auction-theoretic problems that have to be solved. Our proposal specifies a single network entry point with auctioned access. In practice, networks have multiple gateways, each subject to differing states of congestion. Should a smart market be located in a single, central hub, with current prices continuously transmitted to the many gateways? Or should a set of simultaneous auctions operate at each gateway? How much coordination should there be between the separate auctions? All of these questions need not only theoretical models, but also empirical work to determine the optimal rate of market-clearing and inter-auction information sharing, given the costs and delays of real-time communication.

Another serious problem for almost any usage pricing scheme is how to correctly determine

---

1 [MacKie-Mason et al. 1995a] and [Murphy and Murphy 1994] describe an alternative congestion pricing scheme that would set prices based on a current measure of congestion in a gateway, then communicate these to the user. The user would then decide how much traffic to send during the current pricing interval. This mechanism is easier to implement, but at least in principle it does not match the efficiency of the smart market.

2 Since most users are willing to tolerate some delay for email, file transfer and so forth, most traffic should be able to go through with acceptable delays at a zero congestion price, but time-critical traffic will typically pay a positive price.

whether sender or receiver should be billed. With telephone calls it is clear that in most cases the originator of a call should pay. However, in a packet network, both "sides'' originate their own packets, and in a connectionless network there is no mechanism for identifying party B's packets that were solicited as responses to a session initiated by party A. Consider a simple example: A major use of the Internet is for file retrieval from public archives. If the originator of each packet were charged for that packet's congestion cost, then the providers of free public goods (the file archives) would pay nearly all of the congestion charges induced by a user's file request.[1] Either the public archive provider would need a billing mechanism to charge requesters for the (ex post) congestion charges, or the network would need to be engineered so that it could bill the correct party. In principle this problem can be solved by schemes like "800'', "900'' and collect phone calls, but the added complexity in a packetized network may make these schemes too costly.

How large would congestion prices be?

Consider the average cost of the NSFNET backbone in 1993: about $\$10^6$ per month, for about $60{,}000 \times 10^6$ packets per month. This implies a cost per packet (on average about 200 bytes) of about 1/600 cents. If there are 20 million users of the NSFNET backbone (10 per host computer), then full cost recovery of the NSFNET subsidy would imply an average monthly bill of about $0.08 per person. If we accept the estimate that the total cost of the U.S. portion of the Internet is about 10 times the NSFNET subsidy, we come up with 50 cents per person per month for full cost recovery. The revenue from congestion fees would presumably be significantly less than this amount.

The average cost of the Internet is so small today because the technology is so efficient: the packet-switching technology allows for very cost-effective use of existing lines and switches. If everyone only sent ASCII email, there would probably never be congestion problems on the Internet. However, new applications are creating huge demands for additional bandwidth. A video e-mail message could easily use 10^4 more bits than a plain text ASCII e-mail with the "same'' information content and providing this amount of incremental bandwidth could be quite expensive. Well-designed congestion prices would not charge everyone the average cost of this incremental bandwidth, but instead charge those users whose demands create the congestion and need for additional capacity.

[1] Public file servers in Chile and New Zealand already face this problem: any packets they send in response to requests from foreign hosts are charged by the network. Network administrators in New Zealand are concerned that this blind charging scheme is stifling the production of information public goods. For now, those public archives that do exist have a sign-on notice pleading with international users to be considerate of the costs they are imposing on the archive providers.

What are the problems associated with Internet accounting?

One of the first necessary steps for implementing usage-based pricing (either for congestion control or multiple service class allocation) is to measure and account for usage. Accounting poses some serious problems. For one thing, packet service is inherently ill-suited to detailed usage accounting, because every packet is independent. As an example, a one-minute phone call in a circuit-switched network requires one accounting entry in the usage database. But in a packet network that one-minute phone call would require around 2500 average-sized packets; complete accounting for every packet would then require about 2500 entries in the database. On the NSFNET alone nearly 60 billion packets are being delivered each month. Maintaining detailed accounting by the packet similar to phone company accounting may be too expensive.

Another accounting problem concerns the granularity of the records. Presumably accounting detail is most useful when it traces traffic to the user. Certainly if the purpose of accounting is to charge prices as incentives, those incentives will be most effective if they affect the person actually making the usage decisions. But the network is at best capable of reliably identifying the originating host computer (just as phone networks only identify the phone number that placed a call, not the caller). Another layer of expensive and complex authorization and accounting software will be required on the host computer in order to track which user accounts are responsible for which packets.[1] Imagine, for instance, trying to account for student e-mail usage at a large public computer cluster.

One interesting approach has been tested and reported by [Edell et al. 1995]. They found that most traffic could be treated as connection-oriented by tracking the setup and tear-down of TCP sessions. On that basis, they were able to collect real-time usage accounting data on two T-1 lines leaving the UC Berkeley campus, and to introduce a pilot billing server. Another tool is NetTraMet, which has been used in New Zealand for several years to do network accounting [Brownlee 1996].

Accounting is more practical and less costly the higher the level of aggregation. For example, the NSFNET collected some information on usage by each of the subnetworks that connect to its backbone (although these data are based on a sample, not an exhaustive accounting for every packet). Whether accounting at lower levels of aggregation is worthwhile is a different question that depends importantly on cost-saving innovations in internetwork accounting methods.

What are some of the economic problems for commerce on the Internet?

Imagine walking into a bookstore, looking up a book, finding it on the shelves, browsing through

[1] Statistical sampling could lower costs substantially, but its acceptability depends on the level at which usage is measured---e.g., user or organization---and on the statistical distribution of demand. For example, strong serial correlation can cause problems.

its neighbors on the shelve, and finally paying for it with a credit card at the counter. None of this required an explicit set of pre-negotiated contracts or complicated protocols. The value of the Internet will be much greater if this kind of "spontaneous commerce'' becomes commonplace. In order for this to become a reality, it will be necessary to design Internet search and discovery tools, browsing tools, and payment mechanisms. Research on all these topics is underway.

However, the Internet environment also offers new challenges. One big one is security: what protocols can ensure that your credit card number or, for that matter, the details of your purchase, remain private and secure?

How does electronic currency work?

What is so hard about electronic currency? After all, debit and teller machine cards are in common use over networks. Credit cards are widely used over the telephone network. It turns out that there are several difficult problems to be solved, though the problems vary with the type of currency under discussion. For example, bank debit cards and automatic teller cards work because they have reliable authentication procedures based on both a physical device and knowledge of a private code. Digital currency over the network is more difficult because it is not possible to install physical devices and protect them from tampering on every workstation. Credit cards over the phone network are relatively secure because phone tapping is difficult and costly, and there is no central database connected to the network that contains all of the voice-provided credit card numbers. When a credit card number is transmitted over the Internet in the clear, however, "sniffing'' it is relatively easy and inexpensive, and following its path may lead to a massive database of valid card numbers.

A variety of schemes are being developed. See [The Information Economy] and the [Telecom Information Directory] Web pages for comprehensive catalogs of different systems and research papers on this topic. Many of these systems use forms of public key cryptography to encrypt payment records. This is relatively straightforward for, say, credit card numbers, but becomes substantially more difficult if you want to ensure anonymity. See [Chaum 92] for a description of one such electronic cash system.

Why are so many different types of electronic currency being developed?

There are many different types of currency in ordinary, non-network use: cash, personal checks, cashier's checks, money orders, credit cards, debit cards, bearer bonds, and so forth. Each of these has different characteristics along a number of dimensions: anonymity, security, acceptability, transactions costs, divisibility, hardware independence, off-line operation, etc. Likewise, for a rich variety of commercial transactions to develop on the Internet, it will be necessary to have a variety of currency types in use.

How will electronic currency affect the money supply, taxes, illicit activity?

There is already a casino on the Internet, as well as some pornography. It is claimed that various hate groups have used email and bulletin boards for correspondence. Private electronic cash transactions on public networks are likely to facilitate tax evasion and illegal transactions (such as narcotics). No one knows how the introduction of electronic cash will affect macroeconomic variables.

How should information services be priced?

Our focus thus far has been on the technology, costs and pricing of network transport. However, most of the value of the network is not in the transport, but in the value of the information being transported. For the full potential of the Internet to be realized it will be necessary to develop methods to charge for the value of information services available on the network.

There are vast troves of high-quality information (and probably equally large troves of dreck) currently available on the Internet, all available as free goods. Historically, there has been a strong base of volunteerism to collect and maintain data, software and other information archives. However, as usage explodes, volunteer providers are learning that they need revenues to cover their costs. And of course, careful researchers may be skeptical about the quality of any information provided for free.

Charging for information resources is quite a difficult problem. A service like Compuserve charges customers by establishing a billing account. This requires that users obtain a password, and that the information provider implement a sophisticated accounting and billing infrastructure. However, one of the advantages of the Internet is that it is so decentralized: information sources are located on thousands of different computers. It would simply be too costly for every information provider to set up an independent billing system and give out separate passwords to each of its registered users. Users could end up with dozens of different authentication mechanisms for different services. Information discovery and retrieval would suffer from the delays in setting up accounts, as well.

A deeper problem for pricing information services is that our traditional pricing schemes are not appropriate. Most pricing is based on the measurement of replications: we pay for each copy of a book, each piece of furniture, and so forth. This usually works because the high cost of replication generally prevents us from avoiding payment. If we buy a table, we generally have to go to the manufacturer to buy one for ourselves; we can't just simply copy yours. With information goods the pricing-by-replication scheme breaks down. This has been a major problem for the software industry: once the sunk costs of software development are invested,

replication costs essentially zero. The same is especially true for any form of information that can be transmitted over the network. Imagine, for example, that copy shops begin to make course packs available electronically. What is to stop a young entrepreneur from buying one copy and selling it at a lower price to everyone else in the class? This is a much greater problem even than that which publishers face from unauthorized photocopying, since the cost of replication is essentially zero.

There is a small literature on the economics of copying that examines some of these issues. However, the same network connections that exacerbate the problems of pricing "information goods'' may also help to solve some of these problems. For example, [Cox 1992] describes the idea of "superdistribution'' of "information objects'' in which accessing a piece of information automatically sends a payment to the provider via the network. However, there are several problems remaining to be solved before such schemes can become widely used.

REGULATION AND PUBLIC POLICY

What does the Internet mean for telecommunications regulation?

The growth of data networks like the Internet are an increasingly important motivation for regulatory reform of telecommunications. A primary principle of the current regulatory structure, for example, is that local phone service is a natural monopoly, and thus must be regulated. However, local phone companies face ever-increasing competition from data network services. For example, the fastest growing component of telephone demand has been for fax transmission, but fax technology is better suited to packet-switching networks than to voice networks, and faxes are increasingly transmitted over the Internet. As integrated services networks emerge, they will provide an alternative for voice calls and video conferencing, as well. This "bypass'' is already occurring in the advanced private networks that many corporations, such as General Electric, are building.

As a result, the trend seems to be toward removing barriers against cross-ownership of local phone and cable TV companies. The regional Bell operating companies have filed a motion to remove the remaining restrictions of the Modified Final Judgment that created them (with the 1984 breakup of AT&T). The White House, Congress, and the FCC are all developing new models of regulation, with a strong bias towards deregulation (for example, see the [Telecom Act of 1996]).

Internet transport itself is currently unregulated. This is consistent with the principal that common carriers are natural monopolies, and must be regulated, but the services provided over those common carriers are not. However, this principal has never been consistently applied to phone companies: the services provided over the phone lines are also regulated. Many public interest groups are now arguing for similar regulatory requirements for the Internet.

One issue is "universal access,'' the assurance of basic service for all citizens at a very low price. But what is "basic service''? Is it merely a data line, or a multimedia integrated services connection? And in an increasingly competitive market for communications services, where should the money to subsidize universal access be raised? High-value uses which traditionally could be charged premium prices by monopoly providers are increasingly subject to competition and bypass.

A related question is whether the government should provide some data network services as public goods. Some initiatives are already underway. For instance, the Clinton administration has required that all published government documents be available in electronic form. Another current debate concerns the appropriate access subsidy for primary and secondary teachers and students.

What are some of the competing visions for the National Information Infrastructure?

There are probably as many visions of the NII as there are nodes on the Internet. But the two broad models are the Internet model ("many to any'') and the cable TV model ("broadcast to couch potatoes''). A well-written discussion of the "Internet model'' vision is available in [National Academy of Sciences 1994]. One critical issue is the amount of bandwidth provided from the home. The Internet model sees bandwidth as being more-or-less symmetric; the cable TV model sees a much more limited outbound bandwidth: essentially enough for home shopping. As one wit has said about interactive TV networks, "how much bandwidth do you need to send 'I want it' to the Home Shopping Network?''

What will be the market structure of the information highway?

If different components of local phone and cable TV networks are deregulated, what degree of competition is likely? Similar questions arise for data networks. For example, a number of observers believe that by ceding backbone transport to commercial providers, the federal government has endorsed above-cost pricing by a small oligopoly of providers. Looking ahead, equilibrium market structures may be quite different for the emerging integrated services networks than they are for the current specialized networks.

One interesting question is the interaction between pricing schemes and market structure. If competing backbones continue to offer only connection pricing, would an entrepreneur be able to skim off high-value users by charging usage prices, but offering more efficient congestion control? Alternatively, would a flat-rate connection price provider be able to undercut usage-price providers, by capturing a large share of low-value "baseload'' customers who prefer to pay for congestion with delay rather than cash? The interaction between pricing and market structure

may have important policy implications, because certain types of pricing may rely on compatibilities between competing networks that will enable efficient accounting and billing. Thus, compatibility or interoperability regulation may be needed, similar to the interconnect rules imposed on regional Bell operating companies.

How will the choice of service architecture affect the network services available?

The architecture of a network can have important implications for the nature of goods available. For instance, the Internet provides access to an incredibly diverse array of information sources, from personal home pages to fully searchable and professionally managed archives. We believe that the salient feature that drives the diversity of the Internet is that the network provides only bit transportation services; it is up to the end hosts to construct higher-level applications on top of this raw transport service. This architecture has the great advantage that it need not be modified as new applications arise, because applications are implemented entirely at the end hosts and no centralized authority needs to approve such implementations. We call such an architecture application-blind.

There are also a wide variety of services available via 900 numbers on the phone network. In this case the network is application-aware (voice telephony circuits) but content-blind. In comparison, the offerings of cable television, which is content-aware, are rather limited in scope. To what extent do these differences reflect the effect of architecture on the provision of content? [MacKie-Mason et al. 1996b] explore this question, focusing on opportunities to price discriminate, service provider liability, the costs of implementing an aware architecture, and the effects of clutter from the availability of too many applications or too much content.

WHAT ARE OTHER IMPORTANT ECONOMIC PROBLEMS FOR THE FUTURE INTERNET?

How will network distribution and electronic publishing affect intellectual property rights?

One immediate challenge for information service provision over public networks is the definition and protection of intellectual property rights. Existing intellectual property law is far from adequate to handle digital materials. The standard motivation for copyright is that it will encourage the creation and distribution of new works. But if copies of digital works can be produced at zero cost and distributed with perfect fidelity, what will this do to the incentive to produce originals? Several writers have suggested that a new conception of the value of intellectual property, and a new focus on the locus of the value-added will be necessary. See, for example, [Barlow 1994], and [Dyson 1995].

What problems will the Internet face in the next 2 years?

We think that the major network service challenge in the next 2 years will be to find ways to support interconnection. The technical problems are relatively straightforward; it's the accounting and economic problems that are tricky. We think it inevitable that a system of settlements will emerge.

What are settlements? When you place a call to Paris, you transit at least 3 telecommunication networks: your local provider, a long distance company, and France Telecom. These companies keep track of calls and make payments to each other based on how much traffic flows in each direction through their networks. There is a similar system in place for post offices.

Some economists have suggested that such a settlement policy will likely arise for the Internet. Since one carrier imposes costs on another by sending it incremental traffic, it seems appropriate that some monetary payments accompany this traffic. Others argue that traffic flows are sufficiently symmetric that a "no settlements'' policy is workable, especially given the nearly-zero incremental cost of transport (as long as capacity is sufficient). Indeed, to date interconnected Internet networks have not used settlements.

Nonetheless, resource usage is not always symmetric, and it appears that the opportunities to free-ride on capacity investments by other network providers are increasing. For example, suppose a new Internet provider hosts a number of World Wide Web servers near a NAP, and then purchases a very short connection to the NAP. Web traffic flows are very asymmetric: a handful of bytes come in from users making requests, and megabytes are sent back out in response. Thus, for the low cost of leasing a short-distance connection to a NAP, a provider could place a huge load onto other networks to distribute to their users, while this provider does not have to deliver much incoming traffic.[1] Other networks provide substantial "transit" between two other networks, and thus do not receive a direct share of the end-user payment for either end.

The new NAPs (funded by NSF) allow for the possibility that interconnected networks will want to implement settlements. The conditions of use for the NAPs explicitly permit settlements, but they must be negotiated independently by the interconnecting networks; as of this writing it appears that none have yet done so. Further, the necessary technical, accounting, and economic infrastructure is not in place.

[1] If this example doesn't seem compelling because the costs are borne by the networks whose users are generating the demand for the large Web traffic flows, then imagine that the free-riding provider instead services junk email servers that send out vast quantities of unsolicited email. Other users don't want to reach these servers, so this network does not have to provide capacity to handle incoming traffic.

What economic problems will the Internet face in the next 3-5 years?

New protocols such as RTP, RSVP, IPv6 and ATM will become more widespread in this timeframe. Such protocols will be better able to deal with integrated services and congestion management. This should allow for new applications such as video-based conferencing and collaboration tools to become widely used. Also we expect to see some progress made on standardizing new tools for information discovery, search and collaboration.

What economic problems will the Internet face in the next 5-10 years?

Once flexible protocols and killer apps are available, users will likely demand considerably more bandwidth. For example, all-optical networks could spring up in high-density areas [Gilder 1992]. But along with this increase in demand for bandwidth will come a recognition of the commodity nature of network transport. The industry will have to find some way to recover fixed costs. One approach is common carriage and regulation, but we hope that less regulated and more competitive solutions can be found.

FURTHER READING

We have written several papers that provide further details on Internet technology, costs, and pricing problems (see, e.g., [MacKie-Mason and Varian 1995b], [MacKie-Mason and Varian 1995a], [MacKie-Mason and Varian 1995c], and [MacKie-Mason and Varian 1995d]). We maintain two large, comprehensive WWW servers containing links to related information. A comprehensive catalog of electronic materials concerning the economics of the Internet can be found at [The Information Economy] site. A comprehensive directory of information available on the Internet concerning telecommunications more broadly is available at the [Telecom Information Directory] site.

REFERENCES

(Baeza-Yates et al. 1993) Ricardo Baeza-Yates, José M. Piquer, and Patricio V. Poblete. The Chilean Internet connection, or, i never promised you a rose garden. In Proc. INET '93, 1993.

(Barlow 1994) John Perry Barlow. The economy of ideas. Wired, 2(3), March 1994. Available from URL: http://ippsweb.ipps.lsa.umich.edu/ipps744/docs/economy-ideas.html.

(Bohn et al. 1993) Roger Bohn, Hans-Werner Braun, Kimberly Claffy, and Stephen Wolff. Mitigating the coming Internet crunch: Multiple service levels via precedence. Technical report, UCSD, San Diego Supercomputer Center, and NSF, 1993. Available from URL: ftp://ftp.sdsc.edu/pub/sdsc/anr/papers/precedence.ps.Z.

(Braun and Claffy 1993) Hans-Werner Braun and K. Claffy. Network analysis in support of

Internet policy requirements. In Proc. INET '93, 1993. Available from URL: ftp://www.sdsc.edu/pub/sdsc/anr/papers/inet93.policy.ps.Z.

(Brownlee 1994) Nevil Brownlee. New Zealand experiences with network traffic charging. ConneXions, 8(12), December 1994. Available from URL: http://www.auckland.ac.nz/net/Accounting/nze.html.

(Brownlee 1996). Nevil Brownlee. New Zealand Experiences with Network Charging. Whinston. A Priority Pricing Approach to Managing Multi-Service Class Networks in Real Time. Journal of Electronic Publishing. Forthcoming in Internet Economics, J. Bailey and L. McKnight, eds., MIT Press. Available from URL: http://www.press.umich.edu:80/jep/works/BrownNewZe.html.

(Carter and Guthrie 1994) Michael Carter and Graeme Guthrie. Pricing Internet: The New Zealand experience. Technical report, University of Canterbury, Christchurch, New Zealand, 1994. Available at URL: ftp://gopher.econ.lsa.umich.edu/pub/Archive/nz-internet-pricing.ps.Z.

(CERFnet) CERFnet. Network Service Provider Interconnections and Exchange Points. Available at URL: http://www.cerf.net/cerfnet/about/interconnects.html.

(Chaum 1992) David Chaum. Achieving Electronic Privacy. Scientific American, August 1992, p. 96-101. Available at URL: A http://www.digicash.com/publish/sciam.html.

(Cocchi et al. 1992) Ron Cocchi, Deborah Estrin, Scott Shenker, and Lixia Zhang. Pricing in computer networks: Motivation, formulation, and example. Technical report, University of Southern California, October 1992. Available from URL: ftp://ftp.parc.xerox.com/pub/net-research/pricing2.ps.Z.

(Cox 1992) Brad Cox. What if there is a silver bullet and the competition gets it first? Journal of Object-oriented Programming, xx, June 1992. Available from URL: http://repository.gmu.edu/bcox/Cox/CoxWhatIfSilverBullet.html.

(Dyson 1995) Esther Dyson. Intellectual property on the net. Wired, 3(7), July 1995. Available from URL: http://www.eff.org/pub/Publications/Esther_Dyson/ip_on_the_net.html.

(Edell et al. 1995) Richard J. Edell, Nick McKeown, and Praveen P. Varaiya. Billing users and pricing for TCP. IEEE Journal on Selected Areas in Communications, September 1995. Available from URL: http://www-path.eecs.berkeley.edu/ edell/papers/Billing/article.ps.

(Fazio 1995) Dennis Fazio. Hang on to your packets: The information superhighway heads to valley fair. Technical report, Minnesota Regional Network, 1995. Available at URL: http://www.mr.net/announcements/valleyfair.html.

(Gilder 1992) George Gilder. The coming of the fibersphere. Forbes ASAP, xx:111--124, December 1992. Available from URL: http://www.seas.upenn.edu/ gaj1/fiber.html.

(Gupta et al. 1994) Alok Gupta, Dale O. Stahl, and Andrew B. Whinston. Managing the Internet as an economic system. Technical report, University of Texas at Austin, July 1994. Available from URL: http://cism.bus.utexas.edu/ravi/pricing.ps.Z.

(Gupta et al. 1996) Alok Gupta, Dale O. Stahl, and Andrew B. Whinston. A Priority Pricing Approach to Managing Multi-Service Class Networks in Real Time. Journal of Electronic Publishing. Forthcoming in Internet Economics, J. Bailey and L. McKnight, eds., MIT Press. Available from URL: http://www.press.umich.edu:80/jep/works/GuptaPrior.html.

(The Information Economy) Hal Varian. Economics and the Internet, a World Wide Web server. 1994. Available from URL: http://www.sims.berkeley.edu/resources/infoecon

(MacKie-Mason and Varian 1995a) Jeffrey K. MacKie-Mason and Hal Varian. Pricing the Internet. In Brian Kahin and James Keller, editors, Public Access to the Internet. Prentice-Hall, Englewood Cliffs, New Jersey, 1995. Available from URL: ftp://gopher.econ.lsa.umich.edu/pub/Papers/Pricing_the_Internet.ps.Z.

(MacKie-Mason and Varian 1995b) Jeffrey K. MacKie-Mason and Hal Varian. Some economics of the Internet. In Werner Sichel, editor, Networks, Infrastructure and the New Task for Regulation. University of Michigan Press, 1995. Available from URL: ftp://gopher.econ.lsa.umich.edu/pub/Papers/Economics_of_Internet.ps.Z.

(MacKie-Mason and Varian 1995c) Jeffrey K. MacKie-Mason and Hal R. Varian. Pricing congestible network resources. IEEE Journal of Selected Areas in Communications, September 1995. Available at URL: http://www.spp.umich.edu/ipps/papers/info-nets/useFAQs/useFAQs.html.

(MacKie-Mason and Varian 1995d) Jeffrey K. MacKie-Mason and Hal R. Varian. Some FAQs about usage-based pricing. Computer Networks and ISDN Systems 28, 1995. Available from URL: http://www.spp.umich.edu/ipps/papers/info-nets/useFAQs/useFAQs.html. Also in Proceedings of WWW '94, Chicago, Illinois, and in Proceedings of the) Association of Research Librarians 1994.

(MacKie-Mason et al. 1996a) Jeffrey K. MacKie-Mason, John Murphy, and Liam Murphy. The role of responsive pricing in the Internet. Journal of Electronic Publishing. Forthcoming in Internet Economics, J. Bailey and L. McKnight, eds., MIT Press. Available from URL: http://www.spp.umich.edu/spp/papers/jmm/respons.pdf.

(MacKie-Mason et al. 1996b) Jeffrey K. MacKie-Mason, Scott Shenker, and Hal Varian. Network architecture and content provision: an economic analysis. Telecommunications Policy, 1996. Available from URL: http://www.spp.umich.edu/spp/papers/jmm/clutter.pdf.

(Markoff 1993) John Markoff. Traffic jams already on the information highway. New York Times, November 3:A1, November 3 1993.

(Merit Architecture) Merit Network, Inc. Overview of the New Networking Architecture. Available at URL: http://www.merit.edu/nsf.architecture/.index.html

(Merit Statistics) Merit Network, Inc. NSFNET Statistics. Available at URL: http://www.merit.edu/nsfnet/statistics/.

(Murphy and Murphy 1994) John Murphy and Liam Murphy. Bandwidth allocation by pricing in ATM networks. In Proc. of IFIP Broadband Communications '94, Paris, France, March

1994. Available from URL: ftp://gopher.econ.lsa.umich.edu/pub/Archive/Murphy_Bandwidth_Pricing.ps.Z.

(National Academy of Sciences 1994) National Academy of Sciences. Realizing the Information Future. National Academy Press, Washington, DC, 1994. Available from URL: http://xerxes.nas.edu:70/0h/nap/online/rtif/.

(Network Wizards) Network Wizards. Internet Domain Survey. Available from URL: http://www.nw.com/zone/WWW/top.html.

(Telecom Act of 1996) Telecommunications Act of 1996, PUBLIC LAW: 104-104. Available from URL: http://thomas.loc.gov/cgi-bin/query/z?c104:s.652.enr:.

(Telecom Information Directory) J. MacKie-Mason. Telecom Information Resources Directory, a World Wide Web server. Located at URL: http://www.spp.umich.edu/telecom-info.html.

(Zhang et al. 1993) L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala. RSVP: A Resource ReSerVation Protocol. IEEE Network Magazine, 1993. Available from URL: ftp://ftp.parc.xerox.com/pub/net-research/rsvp.ps.Z.

Authors' Note

The authors wish to acknowledge support from National Science Foundation grant SBR-9230481. The most current version of this paper will be available for anonymous ftp, gopher, and in HTML format for World Wide Web access at http://www.spp.umich.edu/ipps/papers/info-nets/Economic_FAQs/FAQs/FAQs.html. Jeff MacKie-Mason is a member of the Department of Economics and the School of Information, University of Michigan, Ann Arbor, MI 48109-1220. Hal Varian is the Dean of the School of Information Management and Science, University of California, Berkeley, CA 94720. They can be reached at jmm@umich.edu and hal@sims.berkeley.edu.