Some Economics of the Internet

by

Jeffrey K. MacKie-Mason
Hal R. Varian
University of Michigan

November 1992
Current version: February 17, 1994

Abstract. This paper was prepared for the Tenth Michigan Public Utility Conference at Western Michigan University March 25–27, 1993. We describe the history, technology and cost structure of the Internet. We also describe a possible smart-market mechanism for pricing congestion on the Internet.

Keywords. Networks, Internet, NREN, NII.

Address. Hal R. Varian, Jeffrey K. MacKie-Mason, Department of Economics, University of Michigan, Ann Arbor, MI 48109-1220. E-mail: `jmm@umich.edu`, `halv@umich.edu`.

Some Economics of the Internet

Jeffrey K. MacKie-Mason
Hal R. Varian

The High Performance Computing Act of 1991 established the National Research and Education Network (NREN). The NREN is sometimes thought of as the "successor" to the NSFNET, the so-called backbone of the Internet, and is hoped by some to serve as a model for a future National Public Network. Substantial public and private resources will be invested in the NREN and other high performance networks during the next 5 to 10 years. In this paper we outline the history of the Internet and describe some of the technological and economic issues relating to it. We conclude with a discussion of some pricing models for congestion control on the Internet.

1. A Brief History of the Internet[1]

In the late sixties the Advanced Research Projects Administration (ARPA), a branch of the U.S. Defense Department, developed the ARPAnet as a network linking universities and high-tech defense department contractors. Access to the ARPAnet was generally limited to computer scientists and other technical users.

In the mid-eighties the NSF created six supercomputer centers which it wanted to make widely available to researchers. Initially, NSF relied on the ARPAnet, Bitnet and several direct university links for this purpose, but planned from the beginning to develop a network connecting the centers. The planners of this new network, the NSFNET, designed it to provide connectivity for a wide variety of research and educational uses, not just for the supercomputers.[2]

The NSFNET was conceived as a backbone connecting together a group of regional networks. A university would connect to its regional network, or possibly to a neighbor university that had a path to the regional network. The regional network hooked into a regional supercomputer. All of the supercomputers were connected together by the high-speed NSFNET backbone, and thus the whole network was linked together.

We wish to thank Guy Almes, Eric Aupperle, Paul Green, Mark Knopper, Ken Latta, Dave McQueeny, Jeff Ogden, Chris Parkin and Scott Shenker for helpful discussions, advice and data. MacKie-Mason was in residence at the Department of Economics, University of Oslo when this paper was completed.

[2] See Lynch (1993) for a brief but detailed history of the Internet.

This design was quite successful—so successful that it soon became overloaded. In 1987 the NSF contracted with Merit, the Michigan regional network, to upgrade and manage the NSFNET. Merit, aided by MCI and IBM, significantly enhanced the capabilities of the network. Since 1985, the Internet has grown from about 200 networks to well over 11,000 and from 1,000 hosts to over two million. About 640,000 of these hosts are at educational sites, 520,000 are at commercial sites, and about 220,000 are at government/military sites in the U.S., with most of the other 700,000 hosts elsewhere in the world. NSFNET traffic has grown from 85 million packets in January 1988 to 37 billion packets in September 1993. This is a 435-fold increase in just over five and a half years. The traffic on the network is currently increasing at a rate of 6% a month, or doubling once a year.[3]

The NSFNET was funded by public funds and targeted for scientific and educational uses. NSF's Acceptable Use Policy specifically excluded activities not in support of research or education, and extensive use for private or personal business. This policy raised a number of troublesome issues. For example, should access be made available to commercial entities that wanted to provide for-profit services to academic institutions?

In September of 1990, Merit, IBM and MCI spun off a new not-for-profit corporation, Advanced Network & Services, Inc. (ANS). ANS received $10 million in initial funding from IBM and MCI. One of the main reasons for establishing ANS was to ". . . provide an alternative network that would allow commercial information suppliers to reach the research and educational community without worrying about the usage restrictions of the NSFNET." (Mandelbaum and Mandelbaum (1992), p. 76). In November 1992, the responsibility for managing NSFNET Network Operations was taken over by ANS. Merit, however, retains responsibility for providing NSFNET backbone services.

In 1991 ANS created a for-profit subsidiary, ANS CO+RE Systems, Inc., designed to handle commercial traffic on ANSnet. It seems apparent that the institutional structure is developing in a way that will provide wider access to private and commercial interests. According to the Program Plan for the NREN, "The networks of Stages 2 and 3 will be implemented and operated so that they can become commercialized; industry will then be able to supplant the government in supplying

[3] The compound growth rate in bytes transported has been 5.8% per month from March 1991 to September 1993, and 6.4% per month from September 1992 to September 1993. This probably underestimates growth in Internet usage because traffic on alternative backbone routes has probably been growing faster. Current traffic statistics are available from Merit Network, Inc. They can be accessed by computer by using ftp or Gopher to `nic.merit.edu`.

these network services." Indeed, in December 1992 the NSF announced that it will stop directly funding a general-use Internet backbone, with the transition likely to occur in 1994.

2. Internet Technology and Costs

The Internet is a network of networks that all use connectionless packet-switching communications technology. Even though much of the traffic moves across lines leased from telephone common carriers, the technology is quite different from the switched circuits used for voice telephony. A telephone user dials a number and various switches establish a dedicated path between the caller and the called number. This circuit, with a fixed allocation of network resources, stays open and no other caller can use those resources until the call is terminated. A packet-switching network, by contrast, uses statistical multiplexing to maximize use of the communications lines.[4] Each circuit is simultaneously shared by numerous users, and no single open connection is maintained for a particular communications session: part of the data may go by one route while the rest may take a different route. Because of the differences in technology, pricing models appropriate for voice systems will be inappropriate for data networks.

Packet-switching technology has two major components: packetization and dynamic routing. A data stream from a computer is broken up into small chunks called "packets." The IP (Internet protocol) specifies how to break up a datastream into packets and reassemble it, and also provides the necessary information for various computers on the Internet (the routers) to move the packet to the next link on the way to its final destination.

Packetization allows for the efficient use of expensive communications lines. Consider a typical interactive terminal session to a remote computer. Most of the time the user is thinking. The network is needed only when data are actually being sent. Holding a connection open would waste most of the capacity of the network link. Instead, keystrokes are accumulated until an `<Enter,Transmit>` key is pressed, at which time the entire buffer is put in a packet and sent across the network. The rest of the time the network links are free to be used for transporting packets from other users.

With dynamic routing a packet's path across the network is determined anew for each packet transmitted. Because multiple paths exist between most pairs of network nodes, it is quite possible

[4] "Connection-oriented" packet-switching networks also exist: X.25 and frame relay are examples of such.

that different packets will take different paths through the network.[5]

The postal service is a good metaphor for the technology of the Internet (Krol (1992), pp. 20–23). A sender puts a message into an envelope (packet), and that envelope is routed through a series of postal stations, each determining where to send the envelope on its next hop. No dedicated pipeline is opened end-to-end, and thus there is no guarantee that envelopes will arrive in the sequence they were sent, or follow exactly the same route to get there.

The TCP protocol breaks a user's data stream into packets, and then reassembles them at the other end. Thus, TCP creates a virtual connection, to make the stream of separate packets look like a single session to a user's application. To identify and reassemble packets in the correct order, TCP packets add their own header to the data. The header contains the source and destination ports, the sequence number of the packet, an acknowledgment flag, and so on. The header comprises 20 (or more) bytes of the packet.

Once a packet is built TCP sends it to a router, a computer that is in charge of sending packets on to their next destination. At this point IP tacks on another header (20 or more bytes) containing source and destination addresses and other information needed for routing the packet. The router then calculates the best next link for the packet to traverse towards its destination, and sends it on. The best link may change minute-by-minute, as the network configuration changes.[6] Routes can be recalculated immediately from the routing table if a route fails. The routing table in a switch is updated approximately continuously.

The data in a packet may be 1500 bytes or so. Recently the average packet on NSFNET carries about 200 bytes of data (packet size has been steadily increasing). On top of these 200 bytes the TCP/IP headers add about 40; thus about 17% of the traffic carried on the Internet is simply header information.

Over the past 5 years, the speed of the NSFNET backbone has grown from 56 Kbps to 45

[5] Dynamic routing contributes to the efficient use of the communications lines, because routing can be adjusted to balance load across the network. The other main justification for dynamic routing is network reliability, since it gives each packet alternative routes to their destination should some links fail. This was especially important to the military, which funded most of the early TCP/IP research to improve the ARPANET.

[6] Routing is based on a dynamic knowledge of which links are up and a static "cost" assigned to each link. Currently routing does not take congestion into account. Routes can change when hosts are added or deleted from the network (including failures), which happens often with about 1 million hosts and over 11,000 subnetworks.

Mbps ("T-3" service).[7] These lines can move data at a speed of 1,400 pages of text per second; a 20-volume encyclopedia can be sent across the net in half a minute. Many of the regional networks still provide T1 (1.5Mbps) service, but these too, are being upgraded.

The transmission speed of the Internet is remarkably high. We recently tested the transmission delay at various times of day and night for sending a packet to Norway. Each packet traversed 16 links, and thus the IP header had to be read and modified 16 times, and 16 different routers had to calculate the best next link for the transmission. Despite the many hops and substantial packetization and routing overhead, the longest delay on one representative weekday was only 0.333 seconds (at 1:10 pm); the shortest delay was 0.174 seconds (at 5:13 pm).[8]

Current Backbone Network Costs

The postal service is a good metaphor for packet-switching technology, but a bad metaphor for the *cost structure* of Internet services. Most of the costs of providing an Internet backbone are more-or-less independent of the level of usage of the network; i.e., most of the costs are fixed costs. If the network is not saturated the incremental cost of sending additional packets is essentially zero.

The NSF currently spends about $11.5 million per year to operate the NSFNET and provides $7 million per year of grants to help operate the regional networks.[9] There is also an NSF grant program to help colleges and universities to connect to the NSFNET. Using the conservative estimate of 1 million hosts and 10 million users, this implies that the NSF subsidy of the Internet is less than $20 per year per host, and less than $2 per year per user.

Total salaries and wages for NSFNET have increased by a little more than one-half (about 68% nominal) over 1988–1991, during a time when the number of packets delivered has increased 128 times.[10] It is hard to calculate total costs because of large in-kind contributions by IBM and

[7] In fact, although the communications lines can transport 45 Mbps, the current network routers can support only 22.5 Mbps service. "Kbps" is thousand (kilo) bits per second; "Mbps" is million (mega) bits per second.

[8] While preparing the final manuscript we repeated our delay experiment for 20 days in October–November, 1993. The range in delay times between Ann Arbor and Norway was then 0.153 seconds and 0.303 seconds.

[9] The regional network providers generally set their charges to recover the remainder of their costs, but there is also some subsidization from state governments at the regional level.

[10] Since packet size has been slowly increasing, the amount of data transported has increased even more.

MCI during the initial years of the NSFNET project, but it appears that total costs for the 128-fold increase in packets have increased by a factor of about 3.2.

Two components dominate the costs of providing a backbone network: communications lines and routers. Lease payments for lines and routers accounted for nearly 80% of the 1992 NSFNET costs. The only other significant cost is for the Network Operations Center (NOC), which accounts for roughly 7% of total cost.[11] In our discussion we focus only on the costs of lines and routers.

We have estimated costs for the network backbone as of 1992–93.[12] A T-3 (45 Mbps) trunk line running 300 miles between two metropolitan central stations can be leased for about $32,000 per month. The cost to purchase a router capable of managing a T-3 line is approximately $100,000, including operating and service costs. Assuming 50 month amortization at a nominal 10% rate yields a rental cost of about $4900 per month for the router.

Table 1.
Communications and Router Costs
(Nominal $ per million bits)[1]

Year	Communications	Routers	Design Throughput
1960	1.00		2.4 kbps
1962		10.00*	
1963	0.42		40.8 kbps
1964	0.34		50.0 kbps
1967	0.33		50.0 kbps
1970		0.168	
1971		0.102	
1974	0.11	0.026	56.0 kbps
1992	0.00094	0.00007	45 mbps

Notes: 1. Costs are based on sending one million bits of data approximately 1200 miles on a path that traverses five routers.
Sources: 1960–74 from Roberts (1974). 1992 calculated by the authors using data provided by Merit Network, Inc.

[11] A NOC monitors traffic flow at all nodes in the network and troubleshoots problems.

[12] We estimated costs for the network backbone only, defined to be links between common carrier Points of Presence (POPs) and the routers that manage those links. We did not estimate the costs for the feeder lines to the mid-level or regional networks where the data packets usually enter and leave the backbone, nor for the terminal costs of setting up the packets or tearing them apart at the destination.

The costs of both communications and switching have been dropping rapidly for over three decades. In the 1960s, digital computer switching was more expensive (on a per packet basis) than communications (Roberts (1974)), but switching has become substantially cheaper since then.[13] We have estimated the 1992 costs for transporting 1 million bits of data through the NSFNET backbone and compare these to estimates for earlier years in Table 1. As can be seen in 1992 the line cost is about eight times as large as the cost of routers.

The topology of the NSFNET backbone directly reflects the cost structure: lots of cheap routers are used to manage a limited number of expensive lines. We illustrate a portion of the network in Figure 1. Each of the numbered squares is an RS6000 router; the numbers listed beside a router are links to regional networks. Notice that in general any packet coming on to the backbone has to move through two separate routers at the entry and exit node. For example, a message we send from the University of Michigan to a scientist at Bell Laboratories will traverse link 131 to Cleveland, where it passes through two routers (41 and 40). The packet goes to New York, where it again moves through two routers (32 and 33) before leaving the backbone on link 137 to the JVNCnet regional network that Bell Labs is attached to. Two T-3 communications links are navigated using four routers.

Technological and Cost Trends

The decline in both communications link and switching costs has been exponential at about 30% per year (see the semi-log plot in Figure 2). But more interesting than the rapid decline in costs is the change from expensive routers to expensive transmission links. Indeed, it was the crossover around 1970 (Figure 2) that created a role for packet-switching networks. When lines were cheap relative to switches it made sense to have many lines feed into relatively few switches, and to open an end-to-end circuit for each connection. In that way, each connection wastes transmission capacity (lines are held open whether data is flowing or not) but economizes on switching (one set-up per connection).

When switches became cheaper than lines the network is more efficient if data streams are broken into small packets and sent out piecemeal, allowing the packets of many users to share a

[13] There is some mismatch between line and router capabilities. In the current design for a T-3 backbone, routers are relatively inexpensive compared to lines. However, in terms of performance, the switches are the bottleneck, and are likely to remain so for some time.

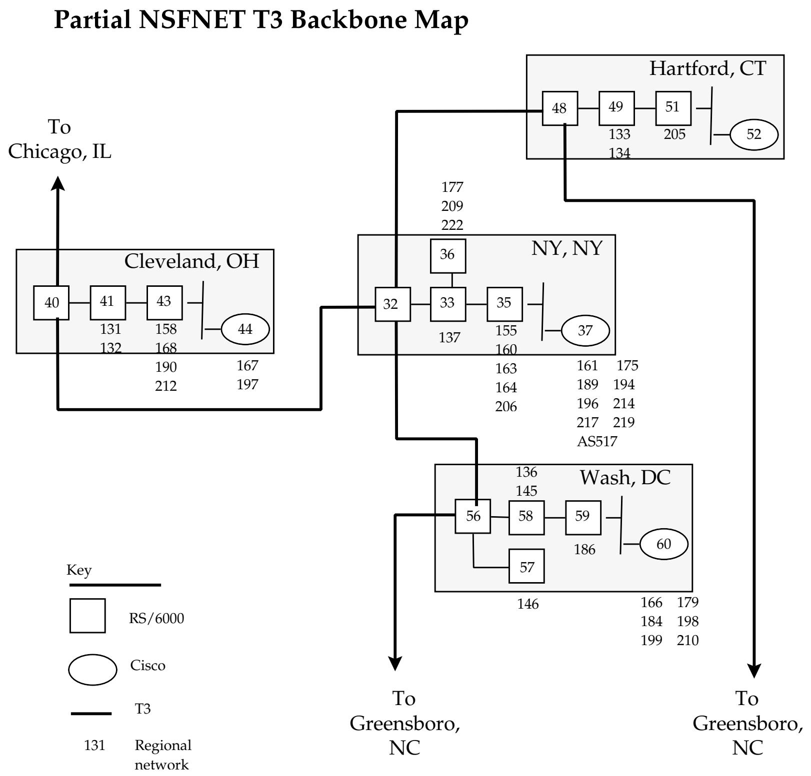


Figure 1. Network Topology Fragment

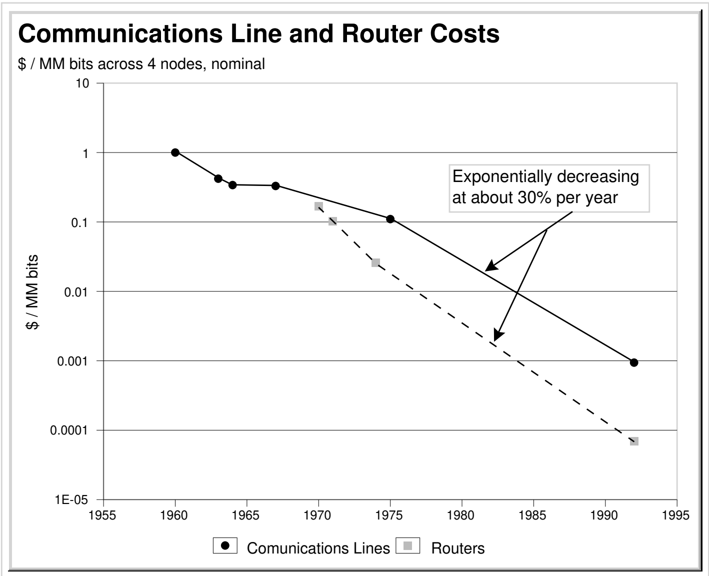


Figure 2. Trends in costs for communications links and routers.

single line. Each packet must be examined at each switch along the way to determine its type and destination, but this uses the relatively cheap switch capacity. The gain is that when one source is quiet, packets from other sources use the same (relatively expensive) lines.

Although the same reversal in switch and line costs occurred for voice networks, circuit-switching is still the norm for voice. Voice is not well-suited for packet-switching because of variation in delivery delays, packet loss, and packet ordering.[14] Voice customers will not tolerate these delays in transmission (although some packetized voice applications are beginning to emerge as transmission speed and reliability increases, see (Anonymous (1986))).[15]

Future Technologies

Packet-switching is not the most efficient technology for all data communications. As we mentioned above, about 17% of the typical packet is overhead (the TCP and IP headers). Since the scarce resource is bandwidth, this overhead is costly. Further, every packet from a data stream must be individually routed through many nodes (12 seems to be typical for a transmission within the U.S.): each node must read the IP header on each packet, then do a routing calculation. Transferring a modest 3 megabyte data file will require around 6,000 packets, each of which must be individually routed through a dozen or so switches.[16] Since a file transfer is a single burst of demand there may be little gain from packetization to share the communications line; for some large file transfers (or perhaps real-time audio and video transmissions) it may be more efficient to use connection-oriented systems rather than switched packets.[17]

Packetization and connection-oriented transport merge in *Asynchronous Transfer Mode* (ATM) which is gaining wide acceptance as the next major link layer technology.[18] ATM does not

[14] Our tests found packet delays ranging between 156 msec and 425 msec on a trans-Atlantic route (N=2487 traces, standard deviation = 24.6 msec). Delays were far more variable to a Nova Scotia site: the standard deviation was 340.5 msec when the mean delay was only 226.2 msec (N=2467); the maximum delay was 4878 msec.

[15] The reversal in link and switch costs *has* had a profound effect on voice networks. Indeed, Peter Huber has argued that this reversal made inevitable the breakup of ATT (Huber (1987)). He describes the transformation of the network from one with long lines all going into a few central offices into a web of many switches with short lines interconnecting them so that each call could follow the best path to its destination.

[16] The average packet size is 350 bytes for FTP file transfers, but for large files the packets will be about 500 bytes each. The header overhead for this transfer would be about 8%.

[17] If there is a slower-speed link on the file transfer path—say 56 kbps—then higher speed links (T-1 or T-3) on the path will have idle capacity that could be utilized if the network is packetized rather than connection-oriented.

[18] The link layer is another layer underneath TCP/IP that handles the transmission and physical congestion control for packets. Current examples of such technologies are Ethernet, FDDI and Frame Relay. The link layer can carry "anyone's" packets; e.g., TCP/IP packets, AppleTalk packets, or Novell Netware packets. Using the postal service analogy, the TCP/IP layer handles "get the mail from New York to Washington; the link layer specifies "mail from NY to DC should be packed

eliminate TCP/IP packetization and thus does not reduce that source of overhead; indeed, ATM includes a 5-byte header in each 53-byte cell, imposing its own 10% overhead.[19] However, ATM opens end-to-end connections, economizing on routing computations, and possibly reducing the overhead from network layer packet headers. Networks currently under development offer speeds of 155 and 622 Mbps (3.4 to 13.8 times faster than the current T-3 lines used by NSFNET). At those speeds ATM networks are expected to carry both voice and data simultaneously. A related alternative is Switched Multi-Megabit Data Service (SMDS) (Cavanaugh and Salo (1992)).

ATM is promising, but we may need radically new technologies very soon. Current networks are meshes of optic fiber connected with electronic switches that must convert light into electronic signals and back again. We are nearing the physical limits on the throughput of electronic switches. All-optical networks may be the answer to this looming bottleneck.

The NSFNET backbone is already using fiber optic lines. A single fiber strand can support one thousand Gbps (gigabits), or about 22,000 times as much traffic as the current T-3 data rate. To give some sense of the astonishing capacity of fiber optics, a single fiber thread could carry all of the phone network traffic in the United States, even during the peak hour of heaviest calling, on Mother's Day (Green (1991)). Yet a typical fiber bundle has 25 to 50 threads (McGarty (1992)), and the telephone companies have already laid some two million miles of fiber optic bundles (each being used at no more than 1/22,000th of capacity) (Green (1991)).

Thus, although switches are cheaper than lines at the rates that current technology can drive fiber communications, in fact we should expect communications bandwidth to be much cheaper than switching before long. Indeed, it is already an electronic bottleneck that is holding us back from realizing the seemingly limitless capacity of fiber. When capacity is plentiful networks will use vast amounts of cheap bandwidth to avoid using expensive switches.

"All-optical" networks may be the way to avoid electronic switches. In an all-optical network data is broadcast rather than directed to a specific destination by switches, and the recipient tunes

in shipping containers and loaded onto a semi-trailer bound for DC." In fact, ATM does not fit neatly into the standard OSI seven-layer network model, because it embeds routing technology, which is traditionally the responsibility of the "network" layer, on top of the link layer.

[19] ATM in fact adds even more overhead, depending on the ATM Adaptation Layer (AAL) protocol that is used to convert a stream of packets into cells and back again. AAL 3/4, originally designed to carry data packets such as TCP/IP, adds another 4 bytes of overhead per cell, for a total of 17% per cell, plus 8 or more bytes per packet (there will typically be many cells per packet). A more efficient protocol, AAL 5, has been developed which adds nearly zero additional overhead per cell beyond the 5-byte ATM header, although it still adds 8 bytes per packet.

in to the correct frequency to extract the intended signal. A fully-functional all-optical network has been created by Paul Green and his colleagues at IBM. His Rainbow I network connects 32 computers at speeds of 300 megabits per second, or a total bandwidth of 9.6 gigabits—200 times as much as the T-3-based NSFNET backbone (Green (1992)).

Despite their promise, all-optical networks will not soon eradicate the problem of congestion. Limitations on the number of available optical broadcast frequencies suggest that subnetworks will be limited to about 1000 nodes, at least in the foreseeable future (Green (1991), Green (1992)). Thus, for an internet of networks it will be necessary to pass traffic between optical subnetworks. The technologies for this are much further from realization and will likely create a congested bottleneck. Thus, although the physical nature of congestion may change, we see a persistent long-term need for access pricing to allocate congested resources.

Summary

We draw a few general guidelines for pricing packet-switching backbones from our review of costs. The physical marginal cost of sending a packet, for a given line and router capacity, is essentially zero. Of course, if the network is congested, there is a social cost of sending a new packet in that response time for other users will deteriorate.

The fixed costs of a backbone network (about $14 million per year for NSFNET at present) are dominated by the costs of links and routers, or roughly speaking, the cost of bandwidth (the diameter of the pipe). Rational pricing, then, should focus on the long-run incremental costs of bandwidth and the short-run social costs of congestion. More bandwidth is needed when the network gets congested (as indicated by unacceptable transmission delays). A desirable pricing structure is one that allocates congested bandwidth and sends appropriate signals to users and network operators about the need for expansion in capacity.

3. Congestion problems

Another aspect of cost of the Internet is congestion cost. Although congestion costs are not paid for by the *providers* of network services, they are paid for by the *users* of the service. Time spent by users waiting for a file transfer is a social cost, and should be recognized as such in any economic accounting.

The Internet experienced severe congestion problems in 1987. Even now congestion problems are relatively common in parts of the Internet (although not currently on the T-3 NSFNET backbone). According to Kahin (1992): "problems arise when prolonged or simultaneous high-end uses start degrading service for thousands of ordinary users. In fact, the growth of high-end use strains the inherent adaptability of the network as a common channel" (page 11). It is apparent that contemplated uses, such as real-time video and audio transmission, would lead to substantial increases in the demand for bandwidth and that congestion problems will only get worse in the future unless there is substantial increase in bandwidth:

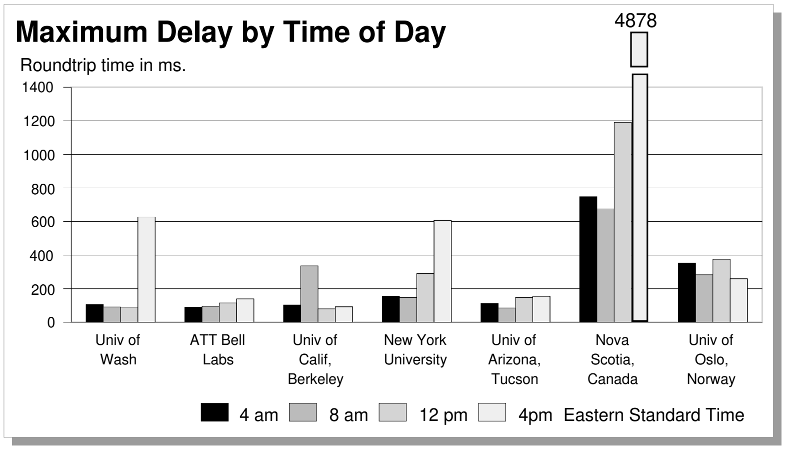
> If a single remote visualization process were to produce 100 Mbps bursts, it would take only a handful of users on the national network to generate over 1Gbps load. As the remote visualization services move from three dimensions to [animation] the single-user bursts will increase to several hundred Mbps . . . Only for periods of tens of minutes to several hours over a 24-hour period are the high-end requirements seen on the network. With these applications, however, network load can jump from average to peak instantaneously." Smarr and Catlett (1992), page 167.

There are cases where this has happened. For example during the weeks of November 9 and 16, 1992, some packet audio/visual broadcasts caused severe delay problems, especially at heavily-used gateways to the Internet NSFNET, and in several mid-level networks.

To investigate the nature of congestion on the Internet we timed the delay in delivering packets to seven different sites around the world. We ran our test hourly for 37 days during February and March, 1993. Deliveries can be delayed for a number of reasons other than congestion-induced bottlenecks. For example, if a router fails then packets must be resent by a different route. However, in a multiply-connected network, the speed of rerouting and delivery of failed packets measures one aspect of congestion, or the scarcity of the network's delivery bandwidth.

Our results are summarized in Figure 3 and Figure 4; we present only the results from four of the 24 hourly probes. Figure 3 shows the median and maximum delivery delays by time of day. Average delays are not always proportional to distance: the delay from Michigan to New York University was generally longer than to Berkeley, and delays from Michigan to Nova Scotia, Canada, were often longer than to Oslo, Norway.

There is substantial variability in Internet delays. For example, the maximum and median delays in Figure 3 are quite different by time of day. There appears to be a large 4pm peak problem on the east coast for packets to New York and Nova Scotia, but much less for ATT Bell Labs



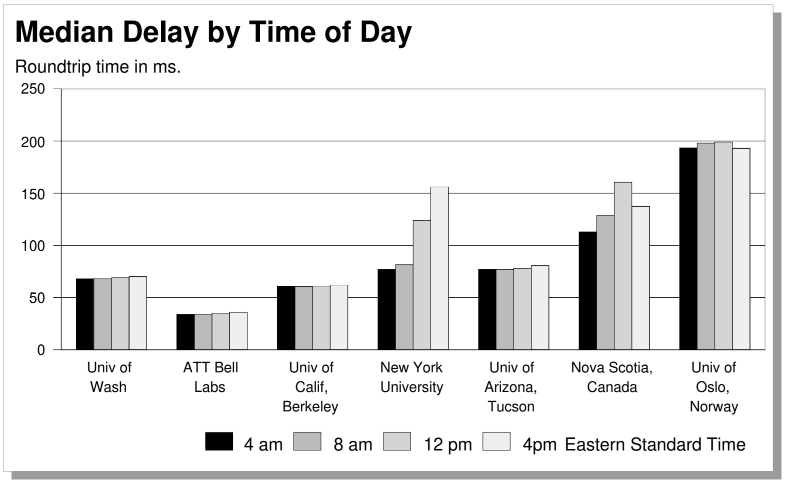


Figure 3. Maximum and Median Transmission Delays on the Internet

(in New Jersey).[20] The time-of-day variation is also evident in Figure 5, borrowed from (Claffy, Polyzos, and Braun (1992)).[21]

Figure 4 shows the standard deviation of delays by time of day for each destination. The delays to Canada are extraordinarily variable, yet the delays to Oslo have no more variability than does transmission to New Jersey (ATT). Variability in delays itself fluctuates widely across times of day, as we would expect in a system with bursty traffic, but follows no obvious pattern.

According to Kleinrock (1992), "One of the least understood aspects of today's networking

[20] The high maximum delay for the University of Washington at 4pm is correct, but appears to be aberrant. The maximum delay was 627 msec; the next two highest delays (in a sample of over 2400) were about 250 msecs each. After dropping this extreme outlier, the University of Washington looks just like UC Berkeley.

[21] Note that the Claffy et al. data were for the old, congested T-1 network. We reproduce their figure to illustrate the time-of-day variation in usage; the actual levels of link utilization are generally much lower in the current T-3 backbone.

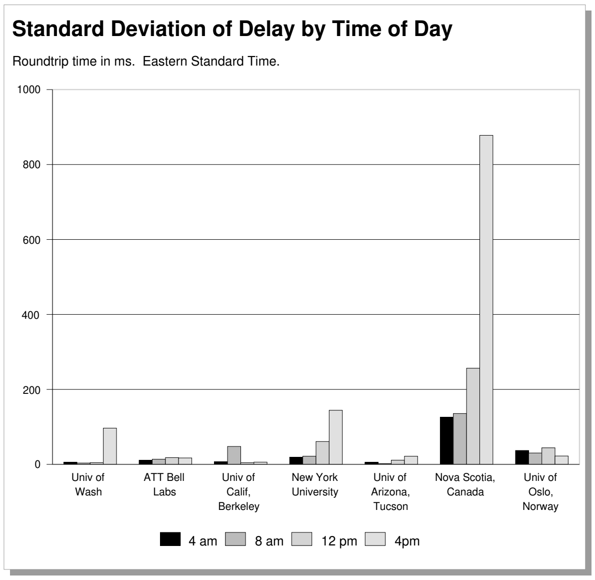


Figure 4. Variability in Internet Transmission Delays

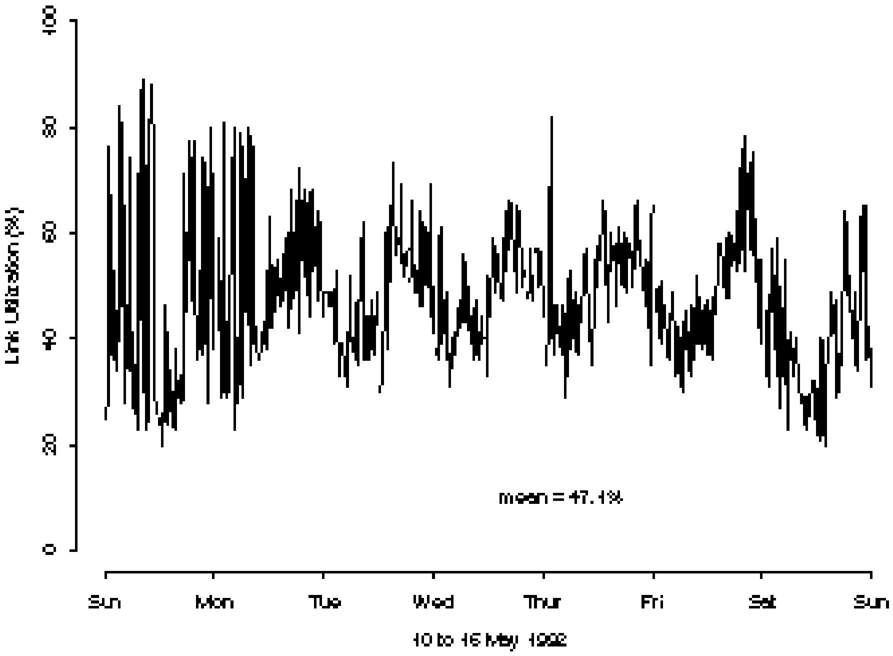


Figure 5. Utilization of Most Heavily Used Link in Each Fifteen Minute Interval (Claffy et al. (1992))

technology is that of network control, which entails congestion control, routing control, and bandwidth access and allocation." We expect that if access to Internet bandwidth continues to be provided at a zero cost there will inevitably be congestion. Essentially, this is the classic problem of the commons: unless the congestion externality is priced, there will inevitably be inefficient use of the common resource. As long as users face a zero price for access, they will continue to "overgraze." Hence, it makes sense to consider how networks such as the Internet should be priced.

As far as we can tell this question has received little attention. Gerla and Kleinrock (1988) have considered some engineering aspects of congestion control. Faulhaber (1992) has considered some of the economic issues. He suggests that "transactions among *institutions* are most efficiently based on *capacity per unit time.* We would expect the ANS to charge mid-level networks or institutions a monthly or annual fee that varied with the size of the electronic pipe provided to them. If the cost of providing the pipe to an institution were higher than to a mid-level network . . . the fee would be higher."

Faulhaber's suggestion makes sense for a dedicated line—e.g., a line connecting an institution to the Internet backbone—but it may not be appropriate for charging for backbone traffic itself. The reason is that the bandwidth on the backbone is inherently a shared resource—many packets "compete" for the same bandwidth. There is an overall constraint on capacity, but there are is no such thing as individual capacity level on the backbone.[22]

Although we agree that it is appropriate to charge a flat fee for connection to the network, we also think that it is important to charge on a usage-sensitive basis, at least when the network is congested. After all, during times of congestion the scarce resource is bandwidth for additional packets.[23] The problem with this proposal is the overhead, or, in economics terms, the transactions cost. If one literally charged for each individual packet, it would be extremely costly to maintain adequate records. However, given the astronomical units involved there should be no difficulty in basing charges on a statistical *sample* of the packets sent. Furthermore, accounting can be done in parallel to routing using much less expensive computers.

Conversely when the network is not congested there is very small marginal cost of sending additional packets through the routers. It would therefore be appropriate to charge users a very small price for packets when the system is not congested.

---

[22] Although it may be true that an institution's use of the backbone bandwidth is more-or-less proportional to the bandwidth of its connection to the backbone. That is, the size of an institution's dedicated line to the backbone may be a good signal of its intended usage of the common backbone.

[23] As we have already pointed out the major bottleneck in backbone capacity is not the bandwidth of the medium itself, but the switch technology. We use the term bandwidth to refer to the overall capacity of the network.

There has been some recent work on designing mechanisms for usage accounting on the Internet. As a first attempt, ANS developed a usage sampling and reporting system it called COMBits. COMBits collects aggregate measures of packets and bytes, using a statistical sampling technique.[24] However, COMBits only collects data down to the network-to-network level of source and destination. Thus, the resulting data can only be used to charge at the level of the subnetwork; the local network administrator is responsible for splitting up the bill (Ruth and Mills (1992)).[25] More recently, the Internet Accounting Working Group has published a draft architecture for Internet usage reporting (Internet Accounting: Usage Reporting Architecture, July 9, 1992 draft). Braun and Claffy (1993) describe measurement of Internet traffic patterns by type of application and by international data flows, and discuss some of the accounting issues that need to be solved. We are also undertaking research on methods for reducing accounting costs.

4. Current Pricing Mechanisms

NSFNET, the primary backbone network of the Internet, has been paid for by the NSF, IBM, MCI and the State of Michigan until the present.[26] However, most organizations do not connect directly to the NSFNET. A typical university will connect to its regional mid-level network; the mid-level maintains a connection to the NSFNET. The mid-level networks (and a few alternative backbone networks) charge their customers for access.

There are dozens of companies that offer connections to the Internet. Most large organizations obtain direct connections, which are leased lines that permit unlimited usage subject to the bandwidth of the line. Some customers purchase "dial-up" service which provides an intermittent connection, usually at much lower speeds. We will discuss only direct connections below.

Table 3 summarizes the prices offered to large universities by ten of the major providers for T-1 access (1.5 mbps).[27] There are three major components: an annual access fee, an initial connection

---

[24] See Claffy, Braun, and Polyzos (1993) for a detailed study of sampling techniques for measuring network usage.

[25] COMBits has been plagued by problems and resistance and currently is used by almost none of the mid-level networks.

[26] NSF restricts the use of the backbone to traffic with a research or educational purpose, as defined in the Acceptable Use Policies.

[27] The fees for some providers are dramatically lower due to public subsidies.

fee and in some cases a separate charge for the customer premises equipment (a router to serve as a gateway between the customer network and the Internet provider's network).[28] The current annual total cost per T-1 connection is about $30–35,000.

Table 2.
Representative Prices for T-1 Connection*

		Fee Components		
		Annual Fee	Initial Connection Cost	Customer Premises Equipment
Service Provider	ALTERnet	24,000	8,900	incl.
	ANS	32,000	incl.	incl.
	CERFnet	20,100	3,750	incl.
	CICnet	10,000	15,000	incl.
	JvNCnet	33,165	13,850	incl.
	Michnet	24,000	14,250	incl.
	MIDnet	6,000	15,000	incl.
	NEARnet	30,000	13,500	incl.
	PREPnet	3,720	1,900	not incl.
	SURAnet	25,000	3,500	3,300

Notes:
* Prices as reported by the vendors. These are prices for a large university. There are some variations in the bundle of services provided, so the prices are not strictly comparable.

Source: Complied by Bill Yurcik, NASA/Goddard Space Flight Center, 11/13/92, with corrections by the authors.

All of the providers use the same type of pricing: annual fee for unlimited access, based on the bandwidth of the connection. This is the type of pricing recommended by Faulhaber (1992). However, these pricing schemes provide no incentives to flatten peak demands, nor any mechanism for allocating network bandwidth during periods of congestion. It would be relatively simple for a provider to monitor a customer's usage and bill by the packet or byte. Monitoring requires only that the outgoing packets be counted at a single point: the customer's gateway router.

However, pricing by the packet would not necessarily increase the efficiency of network service provision, because the marginal cost of a packet is nearly zero. As we have shown, the important

[28] Customers will generally also have to pay a monthly "local loop" charge to a telephone company for the line between the customer's site and the Internet provider's "point of presence" (POP), but this charge depends on mileage and will generally be set by the telephone company, not the Internet provider.

scarce resource is bandwidth, and thus efficient prices need to reflect the current state of the network. Neither a flat price per packet nor even time-of-day prices would come very close to efficient pricing.

5. Proposals for pricing the network

We think that it is worthwhile thinking about how an efficient pricing mechanism might work. Obviously, our suggestions must be viewed as extremely tentative. However, we hope that an explicit proposal, such as we describe below, can at least serve as a starting point for further discussion.

We wholeheartedly adopt the viewpoint of Clark (1989) who says "It is useful to think of the interconnected [networks] as a marketplace, in which various services are offered and users select among these services to obtain packet transport." We take this point of view further and examine what kind of pricing policy makes sense in the context of a connectionless, packet-switched network.

There are many aspects of network usage that might be priced. Cocchi, Estrin, Shenker, and Zhang (1992) make this point quite clearly and describe how a general network pricing problem can be formulated and analyzed. However, we will analyze only one particular aspect of the general network pricing problem in this paper: pricing access and usage of the network backbone.

The backbone has a finite capacity. If enough packets are being sent, some will be excluded or dropped This decline in service quality imposes congestion costs on users. How should a pricing mechanism determine who can send packets at a given time?

6. General observations on pricing

Network engineers tend to take the behavior of the network users as fixed, and try to adapt the technology to fit this behavior. Economists tend to take the technology as fixed and design a resource allocation mechanism that adapts the users' behavior to the technological limitations of the network. Obviously these approaches are complementary!

Let us consider some traditional pricing models for network access. One traditional model is zero-priced access. This is commonly used in highway traffic, for example. This has the well-known defect of the problem of the commons—if each user faces a zero price for access, the network resources tend to become congested.

Most common forms of pricing for network access use posted prices: a fixed price schedule for different priorities of access at different times. For example, the post office charges a fixed price for different priorities of delivery service, and telephone companies provide a fixed charge for connections to different locations at different times of day.

The trouble with posted prices is that they are generally not sufficiently flexible to indicate the actual state of the network at a particular time. If, at a point in time, there is unused capacity, it would be efficient to sell access to the network at marginal cost, which is presumably close to zero. Conversely, if the network is at capacity, some users with high willingness-to-pay may be unable to access the network, even though other users with lower willingness-to-pay have access. Pricing by time-of-day helps to achieve an efficient pattern of usage of network capacity, but it is a rather blunt instrument to achieve a fully efficient allocation of network bandwidth.[29]

7. An ideal but impractical solution

An "ideal" model for network access would be a continuous market in network availability. At each point there would be a price for access to the network. Users who were willing to pay the price for delivery of a packet would be given access; users who weren't would be denied access. The price would be set so as to achieve an optimal level of congestion.

How should the access price be determined? One mechanism would be a "Walrasian tatonnement." A tentative access price would be set. Users would examine the access price and see if they wanted access. If the sum of the demands for access exceed the network capacity the price would be adjusted upward, and so on.

The trouble with this scheme is that the user has to observe the current price in order to determine whether or not he wants access. If the time pattern of usage were completely predictable, there would be no problem. However, packet traffic on the Internet is known to be highly "bursty" and unpredictable.

[29] Posted, flat prices have some benefits. First, accounting and billing use resources too, and may be too high to justify. Second, many planning and budget officers want predictable prices so they can authorize fixed funding levels in advance.

8. A smart market

One way to alleviate this problem is to use a "smart market" for setting the price of network access at different priorities.[30] In a smart market users only indicate the *maximum* willingness-to-pay for network access. We will refer to this maximum willingness to pay as the user's "bid" for network access. The router notes the bid attached to each packet and admits all packets with bids greater than some cutoff value.

We depict the determination of the cutoff priority value in Figure 6. The demand curve indicates how many packets there are at each different bid.

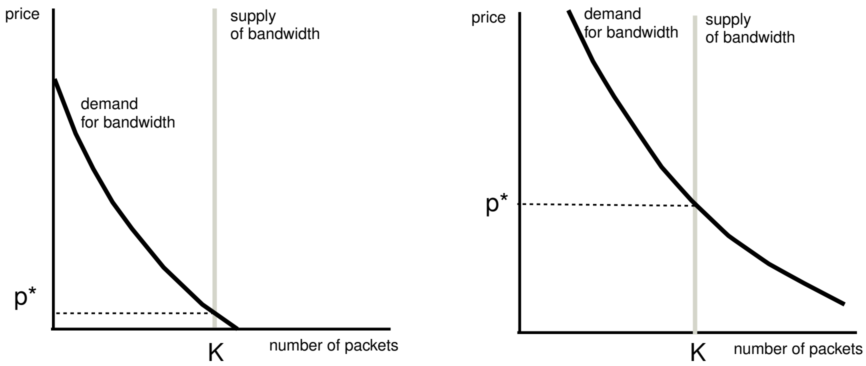


Figure 6. Demand and supply for network bandwidth.

We take the capacity of the network to be fixed, and we indicate it by a vertical line in Figure 6. In the case depicted the demand curve intersects the supply curve at price p^*. Hence, this is the price charged to *all* users—even those who have packets with higher bids.

Note that the bid price can be interpreted as a priority price, since packets with higher bids automatically have higher priority in the sense that they will be admitted before packets with lower bids. This is different from priority-pricing by, say, the post office. In the post-office model you pay for first-class mail even if there is enough excess capacity that second-class mail could move at the same speed. In the smart market described here, a user pays *at most* their bid.

The smart market has many desirable features. It is obvious from the diagram that the outcome is the classic supply-equals-demand level of service. The equilibrium price, at any point in time, is

[30] The term "smart market" seems to be due to Vernon Smith. The version we describe here is a variation on the Vickrey auction.

the bid of the marginal user. Each infra-marginal user is charged this price, so each infra-marginal user gets positive consumer surplus from his purchase.

The major differences from the textbook demand and supply story is that no iteration is needed to determine the market-clearing price—the market is cleared as soon as the users have submitted their bids for access.[31] This mechanism can also be viewed as a Vickrey auction where the n highest bidders gain access at the $n + 1^{st}$ highest price bid.[32]

We have assumed that the bid-price set by the users accurately reflects the true willingness-to-pay. One might well ask whether users have the correct incentives to reveal this value: is there anything to be gained by trying to "fool" the smart market? It turns out that the answer is "no." It can be shown that it is a dominant strategy in the Vickrey auction to bid your true value, so users have no incentive to misprepresent their bids for network access. By the nature of the auction, you are assured that you will never be charged more than this amount and normally you will be charged much less.

9. Remarks about the smart market solution

Here we consider several aspects of using efficient prices for packet access to the Internet.

Who sets the bids?

We expect that choice of bids would be done by three parties: the local administrator who controls access to the net, the user of the computer, and the computer software itself. An organization with limited resources, for example, might choose low bid prices for all sorts of access. This would mean that they may not have access during peak times, but still would have access during off peak periods. Normally, the software program that uses the network would have default values for service—e-mail would be lower than telnet, telnet would be lower than audio, and so on. When special needs arise, the user could override these default values, for example to send an especially urgent e-mail message.

[31] Of course, in real time operation, one would presumably cumulate demand over some time interval. It is an interesting research issue to consider how often the market price should be adjusted. The bursty nature of Internet activity suggests a fairly short time interval. However, if users were charged for packets, it is possible that the bursts would be dampened.

[32] Waldspurger, Hogg, Huberman, Kephart, and Stornetta (1992) describes some (generally positive) experiences in using this kind of "second-bid" auction to allocate network resources. However, they do not examine network access itself, as we are proposing here.

Note that this access control mechanism only guarantees relative priority, not absolute priority. A packet with a high bid is guaranteed access sooner than a low bid, but no absolute guarantees of delivery time can be made.[33] Rejected packets would be bounced back to the sender or buffered until congestion eases, or be routed to a slower network.

Partial congestion

In our discussion we have taken the network capacity to be exogenously given. However, it is easy to extend the mechanism to the case where an additional packet creates congestion for other packets, but does not entirely exclude them. To do this, we simply need use an upward sloping marginal cost/supply curve, rather than a vertical one. We still solve for the same intersection of supply and demand.

Accounting

If the smart market system is used with the sampling system suggested earlier the accounting overhead doesn't have to slow things down much since it can be done in parallel. All the router has to do is to compare the bid of a packet with the current value of the cutoff. The accounting information on every 1000^{th} packet, say, is sent to a dedicated accounting machine that determines the equilibrium access price and records the usage for later billing.[34] Such sampling would require changes in current router technology, however. The NSFNET modified some routers to collect sampled usage data; the cost of the monitoring system is significant.

It is true that routers pose the current technological bottleneck to faster throughput. As a result, some have criticized our smart market idea because it requires extra CPU resources at the routers at the exact same time the routers are a bottleneck. However, the routers are constrained in their ability to keep up with the speed of fiber optic links, while switching and passing large data streams. Accounting and billing should impose much lower CPU requirements that can be handled in parallel by smaller, cheaper CPUs without appreciably slowing the congested routers. That is,

[33] It is hard to see how absolute guarantees *can* be made on a connectionless network. For a discussion of the types of guarantees that might be needed in future network design, see Shenker, Clark, and Zhang (1993).

[34] We don't discuss the mechanics of the billing system here. Obviously, there is a need for COD, third-party pricing, and other similar services.

accounting could be done on a separate machine which would not otherwise be able to reduce the routing bottleneck. Thus, our market may not substantially increase congestion at the router, while it *does* appreciably improve the prioritization of packet delivery during congested times.

Routing

As we have mentioned several times, the Internet is a connectionless network. Each router knows the final destination of a packet, and determines, from its routing tables, what the best way is to get from the current location to the next location. These routing tables are updated continuously to indicate the current topology (but not the congestion) of the network. Routing tables change to reflect failed links and new nodes, but they do not change to reflect congestion on various links of the network. Indeed, there is no standard measurement for congestion available on the current NSFNET T-3 network.

Currently, there is no prioritization of packets: all packets follow the same route at a given time. However, if each packet carried a bid price, as we have suggested, this information could be used to facilitate routing through the Internet. For example, packets with higher bids could take faster routes, while packets with lower bids could be routed through slower links.

The routers could assign access prices to each link in the net, so that only packets that were "willing to pay" for access to that link would be given access. Obviously this description is very incomplete, but it seems likely that having packets bid for access will help to distribute packets through the network in a more efficient way.

Capacity expansion

It is well-known that optimal prices send the correct signals for capacity expansion, at least under constant or decreasing returns to scale. That is, if an optimally priced network generates sufficient revenue to pay the cost of new capacity, then it is appropriate to add that capacity. It appears from our examination of the cost structure of the Internet that constant returns to scale is not a bad approximation, at least for small changes in scale. Hence, the access prices we have described should serve as useful guides for capacity expansion. We have more fully described the role of congestion prices for guiding investment in MacKie-Mason and Varian (1994), which includes a simple analytic model of congestion pricing and capacity expansion.

Distributional aspects

The issue of pricing the Internet is highly politicized. Since usage has been free for many years, there is a large constituency that is quite opposed to paying for it. One nice feature of smart market pricing is that low-priority access to the Internet (such as e-mail) would continue to have a very low cost. Indeed, with relatively minor public subsidies to cover the marginal *resource* costs, it would be possible to have efficient pricing with a price of close to *zero* most of the time, since the network is usually not congested.

If there are several competing carriers, the usual logic of competitive bidding suggests that the price for low-priority packets should approach marginal cost—which, as we have argued, is essentially zero. As a result, in the plan that we have outlined the high priority users would end up paying most of the incremental costs of expanding the Internet.

In any case, our discussion has focused on obtaining an efficient allocation of scarce network resources conditional on the pre-existing distribution of budgetary resources. Nothing about efficient pricing precludes the government from providing cash subsidies for some groups of users to allow them to purchase network access.

10. Role of public and private sector

As we have seen, the Internet is growing at an astonishing rate. The economic setting is also rapidly changing. In December of 1992, the NSF announced that it will stop providing direct operational funding for the ANS T-3 Internet backbone. It is not yet clear when this will actually happen, although the cooperative agreement between NSF and Merit has been extended through April 1994. According to the solicitation for new proposals, the NSF intends to create a new very high speed network to connect the supercomputer centers which would not be used for general purpose traffic. In addition, the NSF would provide funding to regional networks that they could use to pay for access to backbone networks like ANSnet, PSInet, Alternet, and SprintLINK. Meanwhile, the number of commercial vendors of Internet connections and services is rapidly growing, and most of the major telephone companies are getting into the market.

Recent mergers and joint ventures between phone companies, Internet providers, and cable TV operators make it clear that we are moving toward unified networks providing transport for data files, e-mail, voice, video and other multimedia services. The economic issues of the Internet we have discussed are central to the development of markets for integrated services networks. In particular, current providers of access to the Internet generally charge for the "size of the pipe" connecting users to the network. However, the proliferation of services with widely differing requirements for bandwidth, bounded delay, packet ordering and reliability will put enormous strains on backbones or network aggregators that allow unpriced usage of their links.

It is essential that efficient mechanisms for controlling congestion be developed. What has been missing in past efforts to design congestion control are mechanisms for decentralizing the setting of priorities for different demands on network resources. There must be a way to allocate congested bandwidth among millions of users with vastly different uses and different valuations for those uses. We think that usage-sensitive pricing is the most promising mechanism for accomplishing prioritization and socially efficient congestion control.

We have proposed a smart market idea for pricing congestion that has many attractive features. Our proposal is quite preliminary, of course. Several problems must be solved before it would be feasible to implement in a packet-switching network. Current TCP/IP protocols would not support a smart market, but the protocols are evolving. At the same time, technology is changing, so pricing and congestion control mechanisms should be designed to be forward-looking and flexible. For example, the current momentum towards ATM suggests that traffic might be better priced at the ATM link layer rather than at the TCP/IP network layer in the future. ATM may provide a particularly good system in which to embed pricing because it offers connection-oriented service, and out-of-band signalling.[35]

The costs of using a smart market must be carefully weighed against the benefits. One point that we have tried to make clear, however, is that the costs of congestion are real, and are likely to increase dramatically. *Some* mechanism is needed; we hope our proposal stimulates constructive discussion on what mechanisms will be most effective.

There is an important role for the public sector in the evolution of a priced and commercialized Internet. It is very costly to redesign core protocols and implement them throughout the already massive base of installed hardware and software. Designing standards and protocols that can

[35] Users generally want to know the price in advance for a connection, not for a packet, since a given session might send thousands or millions of packets, and prices can change dynamically in a congested network. Out-of-band signalling may help to solve some of the market administration details that are associated with our smart market, or most other pricing mechanisms.

accomodate flexible pricing mechanisms must be done carefully, and work must begin soon. If governments want to see an efficient, competitive and publically beneficial national information infrastructure, they should push now to develop a coherent model for pricing and economic congestion control. A privatized, integrated services Internet will not be viable without such standards.

Glossary[36]

Asynchronous Transfer Mode (ATM)

A method for the dynamic allocation of bandwidth using a fixed- size packet (called a cell). ATM is one type of "fast packet".

backbone

The top level in a hierarchical network. Stub and transit networks which connect to the same backbone are guaranteed to be interconnected. See also: stub network, transit network.

bandwidth

Technically, the difference, in Hertz (Hz), between the highest and lowest frequencies of a transmission channel. However, as typically used, the amount of data that can be sent through a given communications circuit.

Bitnet

An academic computer network that provides interactive electronic mail and file transfer services, using a store-and-forward protocol, based on IBM Network Job Entry protocols. Bitnet-II encapsulates the Bitnet protocol within IP packets and depends on the Internet to route them.

circuit switching

A communications paradigm in which a dedicated communication path is established between two hosts, and on which all packets travel. The telephone system is an example of a circuit switched network.

connectionless

The data communication method in which communication occurs between hosts with no previous setup. Packets between two hosts may take different routes, as each is independent of the other. IP and UDP are connectionless protocols.

Gopher

A distributed information service that makes available hierarchical collections of information across the Internet. Gopher uses a simple protocol that allows a single Gopher client to access information from any accessible Gopher server, providing the user with a single "Gopher space"of information. Public domain versions of the client and server are available.

[36] Most of these definitions are taken from Malkin and Parker (1993).

header

The portion of a packet, preceding the actual data, containing source and destination addresses, and error checking and other fields. A header is also the part of an electronic mail message that precedes the body of a message and contains, among other things, the message originator, date and time.

hop

A term used in routing. A path to a destination on a network is a series of hops, through routers, away from the origin.

host

A computer that allows users to communicate with other host computers on a network. Individual users communicate by using application programs, such as electronic mail, Telnet and FTP.

internet

While an internet is a network, the term "internet"is usually used to refer to a collection of networks interconnected with routers.

Internet

(note the capital "I") The Internet is the largest internet in the world. Is a three level hierarchy composed of backbone networks (e.g., NSFNET, MILNET), mid-level networks, and stub networks. All Internet networks use the IP protocol, with multiple other protocols that run on top of IP.

Internet Protocol (IP)

The Internet Protocol, defined in STD 5, RFC 791, is the network layer for the TCP/IP Protocol Suite. It is a connectionless, best-effort packet switching protocol.

National Research and Education Network (NREN)

The NREN is a US government plan for an interconnected gigabit computer network devoted to Hign Performance Computing and Communications.

packet

The unit of data sent across a network. "Packet"a generic term used to describe unit of data at all levels of the protocol stack, but it is most correctly used to describe application data units.

packet switching

A communications paradigm in which packets (messages) are individually routed between hosts, with no previously established communication path.

protocol

A formal description of message formats and the rules two computers must follow to exchange those messages. Protocols can describe low-level details of machine-to-machine interfaces (e.g., the order in which bits and bytes are sent across a wire) or high-level exchanges between allocation programs (e.g., the way in which two programs transfer a file across the Internet).

route

The path that network traffic takes from its source to its destination. Also, a possible path from a given host to another host or destination.

router

A device which forwards traffic between networks. The forwarding decision is based on network layer information and routing tables, often constructed by routing protocols.

Switched Multimegabit Data Service (SMDS)

An emerging high-speed datagram-based public data network service developed by Bellcore and expected to be widely used by telephone companies as the basis for their data networks.

T1

An AT&T term for a digital carrier facility used to transmit a DS-1 formatted digital signal at 1.544 megabits per second.

T3

A term for a digital carrier facility used to transmit a DS-3 formatted digital signal at 44.746 megabits per second.

Transmission Control Protocol (TCP)

An Internet Standard transport layer protocol defined in STD 7, RFC 793. It is connection-oriented and stream-oriented, as opposed to UDP.

References

Anonymous (1986). Stratacom, inc. introduces 'packetized voice system'. *Communications Week*, *2*.

Braun, H.-W., and Claffy, K. (1993). Network analysis in support of internet policy requirements. Tech. rep., San Diego Supercomputer Center.

Cavanaugh, J. D., and Salo, T. J. (1992). Internetworking with atm wans. Tech. rep., Minnesota Supercomputer Center, Inc.

Claffy, K., Braun, H.-W., and Polyzos, G. (1993). Application of sampling methodologies to wide-area network traffic characterization. Tech. rep. Technical Report CS93-275, UCSD.

Claffy, K. C., Polyzos, G. C., and Braun, H.-W. (1992). Traffic characteristics of the t1 nsfnet backbone. Tech. rep. CS92-252, UCSD. Available via Merit gopher in Introducing the Internet directory.

Clark, D. (1989). Policy routing in internet protocols. Tech. rep. RFC1102, M.I.T. Laboratory for Computer Science.

Cocchi, R., Estrin, D., Shenker, S., and Zhang, L. (1992). Pricing in computer networks: Motivation, formulation, and example. Tech. rep., University of Southern California.

Faulhaber, G. R. (1992). Pricing Internet: The efficient subsidy. In Kahin, B. (Ed.), *Building Information Infrastructure*. McGraw-Hill Primis.

Gerla, M., and Kleinrock, L. (1988). Congestion control in interconnected lans. *IEEE Network*, *2*(1), 72–76.

Green, P. E. (1991). The future of fiber-0ptic computer networks. *IEEE Computer*, *?*, 78–87.

Green, P. E. (1992). An all-optical computer network: Lessons learned. *Network Magazine*, *?*

Huber, P. W. (1987). *The Geodesic Network: 1987 Report on Competition in the Telephone Industry*. U.S. Gov't Printing Office, Washington, DC.

Kahin, B. (1992). Overview: Understanding the NREN. In Kahin, B. (Ed.), *Building Information Infrastructure*. McGraw-Hill Primis, NY.

Kleinrock, L. (1992). Technology issues in the design of NREN. In Kahin, B. (Ed.), *Building Information Infrastructure*. McGraw-Hill Primis.

Krol, E. (1992). *The Whole Internet*. O'Reilly & Associates, Inc., Sebastopol, CA.

Lynch, D. C. (1993). Historical evolution. In *Internet System Handbook*. Addison Wesley, Reading, MA.

MacKie-Mason, J. K., and Varian, H. (1994). Pricing the internet. In Kahin, B., and Keller, J. (Eds.), *Public Access to the Internet*. Unknown, Unknown.

Malkin, G., and Parker, T. L. (1993). Internet users' glossary. Tech. rep., Xylogics, Inc. and University of Texas. Internet Request for Comments 1392.

Mandelbaum, R., and Mandelbaum, P. A. (1992). The strategic future of the mid-level networks. In Kahin, B. (Ed.), *Building Information Infrastructure*. McGraw-Hill Primis.

McGarty, T. P. (1992). Alternative networking architectures: Pricing, policy, and competition. In Kahin, B. (Ed.), *Building Information Infrastructure*. McGraw-Hill Primis.

Roberts, L. G. (1974). Data by the packet. *IEEE Spectrum*, *XX*, 46–51.

Ruth, G., and Mills, C. (1992). Usage-based cost recovery in internetworks. *Business Communications Review*, *xx*, 38–42.

Shenker, S., Clark, D. D., and Zhang, L. (1993). A service model for an integrated services internet. Tech. rep., Xerox PARC and MIT. Internet Draft: draft-shenker-realtime-model-00.ps.

Smarr, L. L., and Catlett, C. E. (1992). Life after Internet: Making room for new applications. In Kahin, B. (Ed.), *Building Information Infrastructure*. McGraw-Hill Primis.

Waldspurger, C. A., Hogg, T., Huberman, B. A., Kephart, J. O., and Stornetta, W. S. (1992). Spawn: A distributed computational economy. *IEEE Transactions on Software Engineering*, *18*(2), 103–117.