CHAPTER 1

Internet Basics

I. A BRIEF HISTORY OF THE INTERNET

The Internet was developed by the U.S. Defense Department's Advanced Research Projects Agency in the early 1969 bearing the name ARPA wide area network (ARPANET). Motivated by the Cold War, the Department of Defense (DoD) was seeking to establish a decentralized communication network which would be more resilient to a bomb attack than the telephone system. To achieve this goal, the DoD would need a robust network capable of functioning when individual nodes were crashed. That is, unlike the telephone system which could be rendered largely useless when a few key carriers were destroyed, the Internet would remain functional. As a result of these objectives, the Internet has the ability to automatically re-route the information it carries, whereas the telephone system does not.

By the late 1970's, several computer networks were using ARPANET; however, the technology did not exist to allow communication between networks. ARPANET could be utilized to establish network communication, but there did not exist a standard technology that would allow different networks to communicate with one another. Therefore, in its early form, ARPANET did not support internetwork communication and was not a 'true' internet, rather, it was a system of independent networks.[1]

Initially, the Internet connected a small group of research centers located at major universities and the DoD. These research centers were mainly involved in military technology projects. Use of ARPANET was soon extended to other major universities. The consequent growth and diversity of the Internet is largely related, directly or indirectly, to the research and applications of universities.

Due to the growth in the demand for Internet usage, by the mid 1980's it was apparent that ARPANET did not have sufficient capacity to meet future needs. In 1989, ARPANET was decommissioned and replaced by a network managed by the National Science Foundation (NSF). The resulting network was known as the NSFNET. Shortly after the development of the NSFNET, it again became clear that capacity was not sufficient to meet growing demand. In addition, the U.S. government realized that it could not continue to fund the Internet indefinitely.[2] Therefore, the government looked to private industry for development and funding. IBM, MERIT, and MCI formed a non-profit corporation to fund research and development of a new wide area network backbone, the ANSNET, the capacity of which was 30 times that of NSFNET. In 1995, MCI developed a new wide area network backbone know as the very high-speed backbone network system, vBNS.[3]

Thus, the Internet has gone through several capacity expansions and technological developments since its inception in the late 1960's. These developments have been driven by the demand for access and bandwidth. In addition to the scope of the Internet's expansion, it is

---

[1] Comer, Douglas E. The INTERNET Book: Everything You Need To Know About Computer Networking And How The Internet Works. 2d ed., (New Jersey: Prentice-Hall, Inc., 1997), 49.

[2] Ibid., 68.

significant that the Internet is attracting the interest and funding of the private sector rather being developed solely by the public sector.

II. NETWORK BASICS

A network is basically a system of interconnected computers. There are primarily two classes of networks, Local Area Networks (LANs) and Wide Area Networks (WANs). Although there are mid-level networks, the distinction is somewhat arbitrary for the purposes of this discussion. Indeed, there are many exceptions to the classification systems use herein.

LANs are the smaller of the two networks; a typical example would be a network of computers within a business organization. In such a setting, the computers are networked so that they can share information without having to transfer it on disk. An additional purpose of LANs is to allow computers to use common resources such as printers and fax machines.

The materials required to set up a LAN are mainly a transmission cable that connects the computers and a circuit board in each computer, which serves as a liaison between the transmission line and the computer. Some of the characteristics of a LAN are: convenience, reliability, speed, low cost, ease of management, and tailorability to various needs.[4] Despite the appeal of these characteristics, there is a large disadvantage; LAN's are not compatible. That is, a LAN cannot be simply connected to another LAN. This incompatibility is largely due to the physical properties of the separate networks such as differences in processing information.

Unlike LANs, Wide Area Networks (WANs) were designed to traverse large geographic regions, e.g., from New York, NY to Chicago, IL. A WAN acts independently of the computers

---

[3] Ibid., 69.
[4] Ibid., 43.

it connects.[5] A WAN is made up of connections maintained by subcomputers. These subcomputers are largely responsible for maintaining the network. Setting up a WAN requires more resources than a LAN does. First, a long-distance transmission line which possesses more bandwidth is needed. Second, computers are needed at several points along the transmission line. Finally, the technology required to transmit information over a WAN is more sophisticated.

A major drawback of a WAN is that it may not be able to execute simple functions. For example, a given WAN may not be able to communicate with a printer. Furthermore, without technology to support interconnection, WANs are not compatible with other WANs nor with LANs.

Midway through ARPANET's existence, there was no mode of standardized communication among computers and networks. Although several LANs and WANs were being used, they were not interconnected due to technological constraints. Each of these independent networks had been designed for various purposes and utilized different modes of operation.

However, the demand to connect these independent networks was ubiquitous. As a result, technologies were developed which allowed the interconnection of networks. Among these technologies, Transmission Control Protocol/Internet Protocol (TCP/IP) became the standard. Although taken for granted today, the development of TCP/IP was a huge breakthrough in utilizing the Internet's technology. TCP/IP allows for the compatibility of LANs and WANs, and allows computers with different operating systems to communicate with one another. Due to the development of TCP/IP the Internet is a network of networks.

III. STRUCTURE OF THE INTERNET

---

5 Ibid., 51.

There are two features which give the Internet its structural integrity. The first is its ability to utilize information in a standardized manner through TCP/IP. The second is its logistical system. An elementary understanding of how the Internet moves information requires the examination of four physical elements which comprise the Internet: modems, service providers, routers, and backbone networks.

The most visible of the four technologies is the modem, which stands for modulator-demodulator. The primary function of the modem is to convert the digital signals leaving the computer into analog signals. Computers utilize digital information, while the line connected to the modem carries analog information. For information traveling from the phone line to the modem, the conversion is from analog to digital.

The Internet service provider if often a less visible part of transmitting information. The service provider is analogous to the local exchange carrier in telephone service. The purpose and organizational structure of the service provider may vary. For example, they may be either for-profit or non-profit. America Online is a for-profit service provider, whereas Illinois State University is a non-profit service provider.

A router is a computer that physically connects networks. A Router receives information, identifies that information, its destination, and assigns it a particular route to its next destination, which may be another router. A single router is connected to several networks.

The backbone network is a type of WAN. Backbone networks traverse large distances, and many small networks feed into them (like rib bones connecting to the backbone). Backbone networks generally carry a greater quantity of information at higher speeds than do the smaller networks.

An example of the Internet using these technologies is as follows: A user dials up the service provider through a modem and dedicated phone line. The service provider then sends the information to a router. The router sends the information through a series of routers, which may or may not use the backbone, until the information reaches the service provider on the receiving end.

IV. CIRCUIT v. PACKET SWITCHING

Although Internet and telephone technologies use similar resources, such as fiber-optic cable, the manner in which these technologies transmit information is different. The distinction between Internet and telephone technology can be more formally examined in the context of **circuit switching** and **packet switching**.

Telephone networks use circuit switching to transmit point-to-point information. Circuit switching is an analog transmission. Circuit switching involves the reservation of a fixed channel for communication; the channel is non-interuptable and uses the same bandwidth regardless of the intensity of use. For example, when placing a phone call, one is guaranteed point-to-point line service. A specific channel, or given amount of bandwidth, is reserved for each phone call. Furthermore, the channel is fixed; the frequency nor the route does not change.

There are two advantages of circuit switching.[6] First, "it enables performance guarantees such as guaranteed maximum delay, which is essential for real-time applications like voice conversations." Maximum delay refers to the amount of time it will take information to arrive at its destination. The second advantage of circuit switching is that accounting for use is much easier, e.g., $.10/minute.

The Internet utilizes packet switching. Packet switching is a form of digital communication. Unlike circuit switching, there is no non-interuptable fixed channel over which information travels. In fact, information from multiple sources may travel over the same bandwidth. Internet information is commingled and independent; therefore, information leaving a particular location at a specific time will arrive by different routes, different order, and different times at its destination.

Utilizing bandwidth in this manner is accomplished by a computer's networking software breaking a transmission into smaller bundles of data called packets. Each packet is identified by a header. The header contains information regarding the packet's source and destination. The rest of the packet contains data. Packets may vary in size depending on the information they contain.

A major advantage of packet switching over circuit switching is that it allows bandwidth to be utilized more efficiently. This is because packets share the available bandwidth. For example, suppose two computers (A and B) want to send information over a given transmission line. After computer A places a packet on the line, computer B will then be allowed to place a packet on the line, and so on, i.e., packets are switched. Thus packet switching allows computers A and B to use the transmission lines simultaneously. (Actually, the lines are not truly used simultaneously; one packet cannot take up the same identical space as another, but communication is more simultaneous than if computer B were required to wait until computer A's entire transmission was sent.)

Another advantage of packet switching is that it does not involve a fixed transmission path. Packet switching allows for packets to travel the path of least resistance in order to avoid congestion. Packet switching uses a system of routers to transmit packets to their destination.

---

[6] Jeffrey K. MacKie-Mason and Hal R. Varian. "Economic FAQ's About the Internet." In Internet

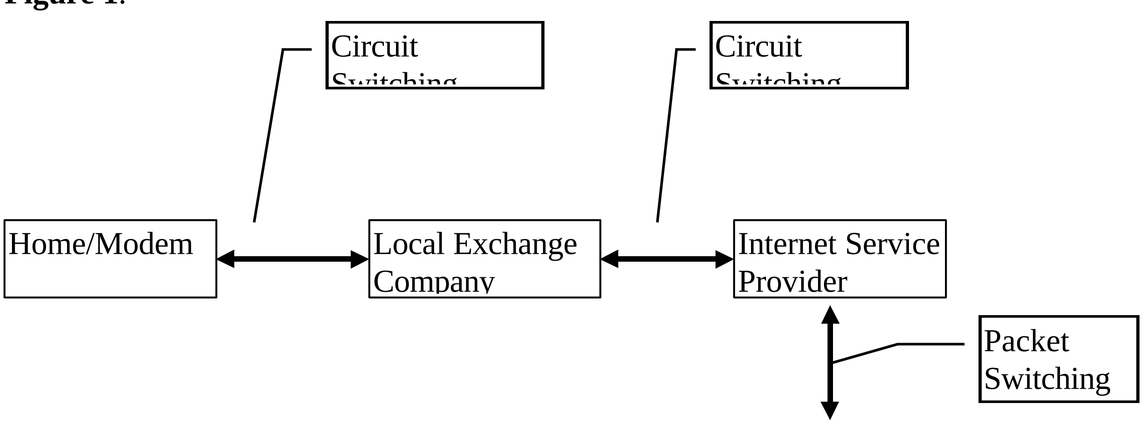
Because packets are independent, the initial router may send different packets to different routers. Therefore, ten packets from a single source may arrive at their destination over ten different routes.

A. Circuit and Packet Switching: A Network Example

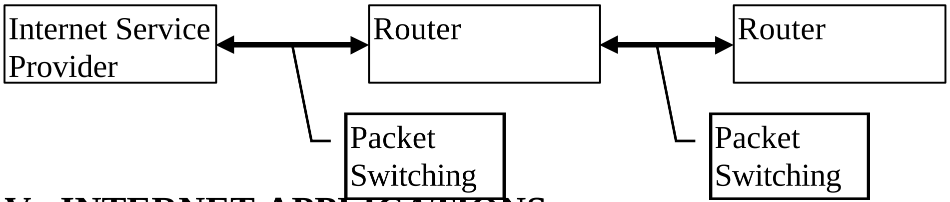
To further understand circuit switching and packet switching an example of the flow of information may be useful. Consider a residential user logging-on from a personal computer at home. Figure 1. illustrates where circuit switching and packet switching are used in the transmission. The information leaves the modem over a dedicated phone line. This phone line connects to the local phone company. The local phone company patches the call through to the Internet service provider. From the time the information leaves the modem until it arrives at the Internet service provider, it is circuit switched (the channel is fixed and non-interuptable).

Once the transmission leaves the Internet service provider, it is packet switched. The packets of information travel over a system of routers until they reach the Internet service provider on the receiving end. Upon arrival, the information may either revert back to circuit switching or remain packet switched.

Figure 1.



Economics, edited by Lee W. McKnight and Joseph P. Bailey. (Cambridge MA: MIT Press, 1997) , 33.



V. INTERNET APPLICATIONS

Although the Internet seems to have one function, to transfer information from point to point, it does have several distinct applications. Below is a list and brief description of some of the most common applications.

- *Electronic mail (e-mail)* Involves a sender and receiver(s)

- *Discussion Groups (Usenet and Listserv)* Involves registering user names on a list. Information can be sent to this list for user review. This allows a large number of people to receive information through a bulk transmission rather than sending the information to each individual user. A news group is an example.

- *Remote Log-in (Telnet)* Remote log-in allows a user to access another computer and negotiate it as if it were the user's computer. This requires that the computer being accessed is a time sharing computer. A time sharing computer allows multiple users to use it simultaneously. Furthermore, the number of users does not affect the performance of a particular user. For example, a professor may be able to access a students data file while the student is working on the file from a different computer. Access to mainframe computers from a personal computer often utilize this application.

- *File Transfer (Anonymous FTP)* Allows a user to download data files efficiently. File transfer only involves one user. The user accesses the data file and downloads it.

- *Navigation (Gopher, Archie, Veronica, Jughead, Wais, World Wide Web (//www.))* More commonly known as web browser applications. These are applications which allow the user to access a myriad of host computers quickly and efficiently.

- *Talk* between two persons allows audio messages to be transmitted over the Internet. Audio messages require a large amount of bandwidth. If there is not sufficient bandwidth to transmit the bandwidth, the message will come through in blocks, and will not be continuos. Video may also be transferred. As with audio the same problems will be encountered with respect to insufficient bandwidth. Although audio requires a lot of bandwidth, video requires approximately a 100 times more bandwidth.[7]

---

[7] Comer, 265.

- *Internet Relay Chat* Conference calls. Same as *Talk,* only with multiple users.

Although these applications may use the Internet structure separately, they are often integrated. For example, a browser may be able to perform FTP and e-mail functions. These applications place different demands on the amount of bandwidth required for transmission. For example, a file transfer generally transfers a larger amount of information and than an e-mail message does. Larger transmissions often take longer to transmit and may result in delayed packets. For example, it is common that delays occur during navigation. A delay during a conference call may not be acceptable because real-time transmission is essential to the quality of service experienced.

CHAPTER 2

Internet Economics

I. WHY AN ECONOMIC INTEREST IN THE INTERNET?

Aside from the Internet's commercial and social usefulness, economists have an interest in the Internet as a scarce resource. Because packets are independent, vary in size, do not use a fixed transmission path, and the total amount of bandwidth is fixed; congestion occurs at different locations called bottlenecks.

As fluid flows from a bottle that possesses a neck, it is channeled into a narrower and narrower path as it approaches the opening. A common example of a bottleneck occurs in traffic situations. Suppose two major highways, each with two lanes running in one direction, merge into a joint two-lane highway for a couple of miles before diverging. The joint two-lane section of highway is now carrying the same number of cars and trucks that four lanes previously carried. As a result, traffic becomes tighter and slows down. Bottlenecking also occurs during road construction when one or more lanes is blocked off. In either case, the traffic is forced into a narrower path and moves at a slower rate.

In the context of the Internet, packets flowing from multiple sources may all need access to a given node or fiber-optic cable in order to reach their destination. As a result, the demand for bandwidth may exceed actual bandwidth. This difference between demand and actual availability is know as congestion. When congestion occurs, the rate of transmission is slowed down, and in some cases, packets may be dropped. In either case, performance degradation occurs.

Aside from occurring at different locations, congestion also occurs at different times. These are periods of peak usage, also known as peak demand. Periods of peak demand occur when it is most convenient for users to transmit information. For example, regular working hours are peak periods simply because they are times when people need to communicate with one another and access information. In settings where individual schedules are not so consistent, the peak period may be much different. For example, peak demand at a university may last from 7:00 am until 11:00 p.m.

A. Tragedy of the Commons

Congestion is an example of the Tragedy of the Commons. The tragedy of the commons is a problem which stems from the communal ownership of pasture land that is used to graze livestock.[8] Consider a pasture surrounding an agricultural community. The acreage of the pasture is fixed, and the quality may not improved during the time period in question. All members of the community have access to the pasture, but no individual member owns any of the pasture. Therefore, one member may not exclude another member from grazing livestock. Communal ownership means that all members can graze their livestock without paying a fee for access to, nor usage of, the pasture.

Since access and use of the pasture are free, each member faces an incentive to graze as much livestock on the pasture as time constraints permit. Despite the apparent utopia of communal ownership, there is a tendency to exhaust the pasture's usefulness. At some point,

---

[8] Robert Cooter and Thomas Ulen, LAW AND ECONOMICS (Harper-Collins, 1988), 186.

grazing an additional head of livestock will decrease the amount and quality of the pasture available to all of the current livestock. That is, the average amount of pasture available to each head will no longer meet the nutritional requirement per head. When this occurs, the pasture is considered to be overgrazed. Thus communal ownership of the pasture leads to overgrazing, and in a short period of time the pasture may become barren and unable to serve a useful purpose.

Although access and use of the pasture are free, there is a cost associated with placing an additional head of livestock on the pasture. This cost is not incurred exclusively by any one member, rather it accrues to all members on a proportionate basis. At the point of overgrazing, if member A places more livestock on the pasture, then the value that the remaining members receive from their livestock diminishes. Because each head is no longer consuming its daily requirement, the heartiness of the livestock will begin to diminish. If the owner chooses to consume his livestock, the quality of the steaks will be lower than those of a hearty head. If some of the livestock is sold on the market, the market value and the price received by the owner will also be less. The difference between the market value of a hearty head of livestock and that of a malnourished head of livestock may be thought of as the cost of overgrazing. This cost accrues to all owners and is, therefore, considered a social cost.

The tragedy of the commons is an analogy for Internet access. With respect to the Internet, bandwidth may be thought of as the pasture land. The relevant qualities are even similar. Both the pasture land and bandwidth represent physical space that is required to support a specific function. Like access and use of the pasture, access and use of the Internet are free; therefore, it may be thought of as communal property.

Each head of livestock is analogous to a packet that is placed on the network. Packets demand bandwidth just as a head of livestock demands a portion of the vegetation on the pasture.

Overgrazing the pasture is analogous to congestion on the Internet. As is the case with overgrazing, allowing an additional packet to be placed on the Internet during congested periods costs the user nothing, but it decreases the value of the Internet to all other users. On the Internet, the social cost is one of performance degradation in the form of delayed or dropped packets rather than malnutrition. With demand for access and the number of hosts increasing, congestion could become such a problem that the existing Internet will no longer serve a useful purpose. This was actually the fate of the ARPANET backbone. So many users were logging-on that it became difficult to transmit a simple message.

Presently, the only incentive for a user to curb their use of the Internet is to avoid congestion. For example, a researcher wanting to transfer a large data set using FTP may wait until the period of peak demand has passed to ensure that congestion is not the cause of some of the packets being dropped. Similarly, a user who normally stays up late at night may forego daytime usage for speeder and more reliable off-peak transmission.

B. Avoiding the Tragedy: Managing Congestion

The tragedy of the (Internet) commons may be avoided with the aid of some type of intervention. Ideally, the form of intervention would discourage use during periods of congestion. Although the goal is to alleviate congestion, economists approach the problem differently depending upon the relevant time period: the short run or the long run.

Economically, the **short-run** is an ambiguous and context specific time period requiring a few assumptions. With respect to the Internet, the short run is assumed to be the period of time between the present and future in which the Internet infrastructure is fixed. That is, bandwidth cannot be expanded and technological advances which enhance service do not occur. These assumptions are reasonable given that it may take years to build a new network or develop new technology to improve service. Given fixed resources, congestion cannot be alleviated by adding additional bandwidth. Under these conditions, the short-run objective is one of developing a policy to allocate bandwidth to some users to the exclusion of others so that congestion is alleviated.

The **long-run** assumption is that the infrastructure of the Internet is no longer fixed and may be varied. The long-run method of managing congestion is to expand bandwidth and utilize technology to such an extent that congestion is precluded. Precluding congestion means installing enough bandwidth so that all users may transmit without imposing a social cost on any of the other users. Given this objective, the problem is one of developing a method of financing the Internet's infrastructure while ensuring its integrity.

II. WHAT TYPE OF GOOD IS THE INTERNET?

A. Private vs. Public Goods

Before developing a short-run or long-run method of alleviating congestion, it is important to define what type of economic good the Internet is. Note, in this context, good refers to a tangible item, not an ethical position. The characteristics of a good largely determine the most efficient method of its allocation.

Goods may generally be classified as private goods or public goods. Characteristics of private goods are excludability and depletability.[9] A private good is used by an individual or group to the exclusion of others. Depletability implies that as a good is used, it is also used up. The good is finite. That a good or service is depletable and excludable comes as no surprise, for this is what everyday experience reveals. Private goods are those which are bought and sold on the market everyday.

In contrast to a private good, a public good is nondepleteable and nonexcludable. Nondepletability means that once the good is produced, its supply is continuous and does not diminish no matter how many people use it. Nonexcludability means that once the good is produced access to it cannot be restricted.

Note that the economic definition of a public good may differ from the casual use of the term. In a casual sense, a public good only implies nonexcludability. In a casual sense, a park is a public good; however, according the economic definition it may not be a true public good. All who wish go the park may. However, the quality of the park may diminish, perhaps due to noise and trash, as more and more people show up. A park is nonexcludable, but depletable.

A common example of a public good is national defense. With respect to nondepletability, once the defense infrastructure and protocols are in place, all citizens benefit regardless of how many citizens use it. If the population increases by 1000 citizens, each of the additional citizens will be protected to the same extent that the members of the original population. With respect to nonexcludability, a particular class of citizens may not be excluded from receiving the benefits of national defense. For example, during a war, citizens of Iowa could not be excluded from protection.

---

[9] Martyne M. Hallgren and Alan K. McAdams. "The Economic Efficiency of Internet Public Goods." In Internet Economics, edited by Lee W. McKnight and Joseph P. Baily. (Cambridge MA: MIT Press, 1997), 457.

C. The Short-Run Trade-Off

The short-run definition of the Internet as a good varies depending on whether or not congestion is present. If there is no congestion, then the Internet possesses the characteristics of a public good (it is both nondepletable and nonexcludable). As a public good, the full value of the Internet is realized. All users who wish to access it may and no user suffers performance degradation.

If congestion is present, then the Internet is a public good only in the casual sense. Much like the park, it is nonexcludable but depletable. Access is not restricted, but the quality of service is depleted when congestion begins to occur. If a good is nonexcludable, depletability may occur.

Despite these characteristics, the Internet is more valuable if it is nondepletable because it would not suffer performance degradation. In the short run, the Internet can be transformed into a nondepletable good if it is made excludable. Making the Internet into an excludable good requires some form of external intervention which effectively discourages usage during congested periods. If congestion is alleviated by excluding individuals from access, the remaining users will suffer no performance degradation. Therefore, in the short run, open access may be traded for quality of service.

III. SOCIALLY EFFICIENT PRODUCTION OF PRIVATE AND PUBLIC GOODS

A. Service Provider Costs

Although the current total cost of the Internet is unknown, it is straightforward to examine what type of costs are incurred and how they are allocated in supplying the Internet. Most of the costs for service providers are either **sunk** or **fixed**. Costs are sunk because they do not depend on present or future decision making. The costs are fixed because they do not vary with the number of users that access and use the network.[10]

In addition, there are two main elements of network costs. First is the cost of providing additional network capacity. This cost may vary depending on the extent of additional capacity added. These costs mainly consist of the fiber-optic cable, router(s), and the labor time required to design and construct additional capacity.[11]

Second is the cost of connecting to the Internet. Once capacity is installed the cost for connection and use of the network is negligible, may "not [be] worth charging for," given accounting and billing costs.[12] For example, the cost to the owner of a fiber-optic cable of letting additional packets on the network is zero. The extra use does not add wear and tear to the cable.

B. Criteria for Efficient Pricing

From a social perspective, the efficient price of a good or service is where price is set equal to the additional cost of supplying that good. When this condition is met, the quantity of the good produced is such that society achieves the greatest benefit from its production. The cost of devoting societies resources to supplying the last (marginal) unit of a particular good is equal to the marginal benefit society achieves form it.

1. Private Goods The marginal cost of supplying private goods is (usually) an increasing and positive function of quantity supplied to additional users. Thus, it is economically straightforward to determine what constitutes the efficient production of private goods. With private goods the producer receives the price signal as a willingness to pay for a good or service. If this price is

---

[10] David W. Crawford. "Internet Services: A Market for Bandwidth or Communications." In Internet Economics, edited by Lee W. McKnight and Joseph P. Baily. (Cambridge MA: MIT Press, 1997), 382.

[11] Jeffrey K. MacKie-Mason and Hal R. Varian, 39.

above the producers opportunity cost the producer will supply it. In the end, a supplier will continue to supply a good until the price received for the last unit is equal to the opportunity cost of producing it.

[12] Ibid., 40

2. Public Goods The cost of supplying a public good to additional users is zero. Given the criteria for efficient pricing, a public good should be priced at zero. Since the Internet is a public good in the long run, and is sometimes a public good in the short run, access and use should be priced at zero.

CONCLUSION

It has been shown that the Internet has several economic characteristics. In attributing characteristics to the Internet, it must first be determined if the relevant context is in the short run or the long run. In the long run, the Internet is a true public good, as a result, it is nondepletable and nonexcludable.

In the short run, the Internet may or may not be a public good depending on whether or not congestion is present. If there is no congestion, the Internet is a public good. However, if congestion is present, the Internet may not be defined strictly as a public good or a private good. Absent intervention, the Internet has the characteristics of depletability and nonexcludability. However, the Internet is more valuable if it is nondepletable and excludable. This is the short-run trade-off. This trade-off can be gained by excluding users to such an extent that congestion is not present.

CHAPTER 3

Managing Congestion

In the previous chapter, it was mentioned that the largest problem of offering Internet service (in line with its full economic value) is both a short-run and long-run problem. This is because the Internet has different characteristics in the two periods. Consequently, the problem of congestion must be addressed differently in each of these contexts.

I. THE SHORT RUN PROBLEM

There are two approaches which may be used to allocate congestion in the short run. Both pricing and non-pricing schemes may transform the Internet from being nonexcludable and depletable to being excludable and nondepletable.

A. Pricing Schemes

Economists traditionally advocate pricing to allocate resources. Pricing serves to allocate goods, in this case bandwidth, to its most highly valued use. As an example, consider two individuals about to log-on. If neither individual uses the Internet, no congestion occurs. However, if both users transmit information, congestion will occur. If congestion is present, the transmission will be delayed. Further, suppose that one individual is an investment banker and the other is a six-year old child wanting to play a game.

During the time in question, the investment banker needs access to important financial information that will have a large impact upon himself and his clients, and time is of the essence. Meanwhile, the six-year old is considering logging-on because there is nothing good on television.

In this example it is obvious the banker values the speedy and reliable transmission of an uncongested Internet more than the six-year old. If a price were placed upon the available bandwidth so that one could purchase it, thereby excluding the second, the banker would be willing to pay to have his transmission sent in a timely and reliable manner. Even if the six-year old was also willing to pay, the banker would be willing to outbid the six-year old. In this example, pricing serves to prioritize users according to need. Absent such a mechanism, all transmissions receive the same priority regardless of the purpose they serve.

In addition to the ability to allocate bandwidth to those users who value it most, pricing serves other purposes. First, it (may) recover costs. Second, it can allow users to select among a choice of services, so that users who wish to use more resources can pay accordingly.[13] There are two general types of pricing schemes: those that base the fee on the intensity of use and those that do not.

1. Flat Fee Flat-fee pricing, also known as connection or subscription pricing, is based on access but is independent of the intensity of use. Generally, the user pays a service provider a monthly or hourly fee for Internet access. Flat fees may also be based on the "capacity of the access link,"[14] with larger capacity commanding a larger fee.

Flat fee pricing is the most common type of Internet service pricing to date. This type of scheme is typical in large organizations. The organization pays an annual fee and the cost to the

---

[13] David D. Clark. "Internet Cost Allocation and Pricing." In Internet Economics, edited by Lee W. McKnight and Joseph P. Bailey. (Cambridge MA: MIT Press, 1997), 229.

member users is zero and usage is unlimited. The following points provides some of the benefits associated with flat fee pricing.[15]

- Flat fees are easy for users to understand. Furthermore, the income stream of the provider is smoother This removes uncertainty about future revenue. Less uncertainty makes for easier business planning.
- Flat Fees encourage usage, which (absent congestion) increases customer satisfaction.
- Flat Fees are simpler to administer, monitor, and ultimately bill for.

2. Basic Usage Sensitive Pricing The basic idea behind usage sensitive pricing is that users pay a per unit fee; in this case a price is associated with each packet placed on the network. Therefore, pricing is based upon the intensity of use. Users placing more packets on the Internet will pay more in fees.

Basic usage sensitive pricing can be structured two different ways. First, all packets may be priced the same regardless of the amount of congestion on the network. Second, the price of a packet may vary with the amount of congestion on the network.

Under the first structure, the service provider and the user may simply contract for an agreed upon price per packet placed upon the network. The user's bill would simply be the number of packets (N) multiplied by the price per packet (P).

Under this structure, the user has an incentive to economize on use, whereas under the flat-fee structure no such incentive exists. However, this approach does not target curtailing use during periods of congestion, nor does it encourage use during noncongested periods. Users pay the same price per packet whether or not congestion is present.

---

14 Ibid.

15 Ibid., 230.

This approach may also violate the efficient pricing rule. If users pay a per-packet fee even when congestion is not present, then the Internet is not priced as a public good.

In the long run, usage sensitive pricing that is not sensitive to congestion may retard network investment and technological advance. In the long run, demand is more elastic for frequent users with large bandwidth requirements; therefore, they will seek to avoid the usage fees by investing in alternative information transfer systems. This leaves the service provider left only to capture the revenues from the small users. And it may be the case that the revenues from small user fees would not be enough to cover costs. This insolvency would cause the service providers to raise prices to cover costs. The larger prices would then provide the incentive to seek alternative methods of information transfer.[16]

The second usage sensitive pricing structure is congestion sensitive. As congestion increases, the price of placing an additional packet on the network increases. This type of pricing discourages use during congested periods and encourages use during non-congested periods. As a result, use across time periods is leveled out and bandwidth is used more efficiently. Therefore, pricing bandwidth only during congested periods discourages the social cost associated with congestion.

Despite the theoretical advantages of usage based pricing there are disadvantages associated with it. The first difficulty arises when considering how to monitor and bill for usage on a real-time basis. If proper billing is not implemented, then the costs associated with placing an additional packet on the network are not accounted for. This problem is compounded by the fact that congestion and prices may peak suddenly and unexpectedly. These random peaks create uncertainty and may be difficult for users to respond to. As a result, usage sensitive pricing may not fully serve as a mechanism to allocate scarce bandwidth. It may even be arbitrary in some circumstances. In the end, usage sensitive pricing may only be seen as a barometer of the level of prices expected within a given block of time, e.g., 5 min. or the next hour.

3. Committed Information Rate Pricing Committed information rate (CIR) pricing is a two-part price. The first part of the price is based on the connection (a flat fee). The second part of the price is based upon the guaranteed flow (CIR) to the customer.[17] A CIR represents the amount of information that can be transmitted by any single user assuming all users are using the network. Therefore, the CIR is a "worst-case rate."[18]

There are two problems in using the CIR as a benchmark for pricing. First, it is assumed that all transmissions of "the user (represent) a steady flow, while in practice the traffic is extremely variable or bursty."[19] The second problem is that all users are not using the network simultaneously; therefore, the actual flow rate experienced by a given user will often exceed the CIR.

The two problems of using the CIR as a benchmark for prices translate into improperly structured incentives for usage. Because the CIR is a fictional and not an actual flow rate, it will not account for congestion. Therefore, users demand bandwidth in the same manner as they would under a usage sensitive scheme that is not sensitive to congestion. Furthermore, customers will not judge the quality of service by the CIR but by actual performance.[20] What they get and what the pay for are almost always two different things.

---

16 Ibid., 231.
17 Jeffrey K. MacKie-Mason and Hal R. Varian, 39.
18 David D. Clark, 223.
19 Ibid., 224.
20 Ibid.

CIR price may help the service provider cover its costs while making it easy for users to predict their billing charge. However, CIR pricing does not provide a method of allocating bandwidth during periods of congestion.

4. Expected Capacity Pricing Scheme Using expected capacity as a basis for pricing provides several improvements over flat fee, basic usage sensitive, and CIR pricing. Expected capacity pricing is a form of usage pricing. Rather than prices tracking actual use, prices are based upon the user's expected use. The service provider and the user work together to determine a user's expected usage based upon the user's profile. The profile may take into account such variables as the time of day the user transmits most, the size of transmissions, and the total time per period the user transmits. These variables are then used to predict the user's expected capacity demand.

Once a user and provider contract for a given profile, the number of packets a user sends are monitored at the service provider's gate. If the number of packets fall within the user's expected range they are tagged *in* . If the user transmits more than expected capacity, packets are tagged *out*.

The metering of tags is done by a "traffic meter" which monitors packets using a system called "token bucket metering."[21] Under this scheme, the meter is thought of as containing a bucket of tokens, the number of which corresponds to expected usage. For example, the number of tokens may correspond to the number of packets or amount of data specified in the user's expected capacity. The tokens in the bucket are replaced at a constant rate so that at any given time, the amount of tokens in the bucket are representative of expected usage.

[21] Ibid., 234.

As packets arrive at the meter, two things happen. First, the traffic meter tags the packets. If there are enough tokens in the bucket to match the data in the packet, the packet is tagged in. If there are not enough tokens in the bucket, the packets are tagged out. Second, the amount of tokens required to tag the packet are subtracted from the bucket.

Under the expected capacity pricing scheme users have an incentive to stay within expected capacity. If the user purchases expected capacity lower than their actual usage, then packets will be delayed or dropped due to congestion. Therefore, in attempting to save money by buying less expected capacity, the user will pay with time and frustration.

Expected capacity has several distinct advantages over the previous pricing methods. First, it does not violate the efficient pricing rule. The user is not penalized for transmitting during periods of congestion. If there is no congestion, the packet does not get dropped. However, when there is congestion the user pays an additional cost in the form performance degradation.

Second, it tracks the costs of the service provider. This is because "the provider must provision to carry the expected capacity of his subscribers during normal busy periods. . . "[22] Whereas basic usage pricing provides an incentive for large users to invest in alternative methods of transmission; the expected capacity system provides no such incentive. In fact, so long as the expected capacity is priced in accordance with the service providers costs, large users would find it difficult to develop a more cost effective method of transmission.

Third, an expected pricing scheme is easier to implement than basic usage pricing. Packets are tagged as soon as they leave the service provider as *in* or *out*. The routers identify the tag and let the packet pass or drop it. Because the meter is at the gate of the service provider, the service provider can monitor the meter easily. This ease of monitoring translates into accurate billing.

Fourth, it is easier for the user to understand than basic usage pricing and CIR pricing. It may be difficult to explain to all customers the complexity involved in pricing individual packets placed on the network. Unlike CIR pricing, the user designs his expected usage profile. This allows the user to have a good sense of the impact of his transmission upon the network.

Finally, the service provider can tailor various profiles to meet the needs of the user. As an example of profile tailoring, "a user who wants to send large files, but only very occasionally, could purchase a very small average [token replacement] rate, but a very large token bucket."[23] Such tailorability may be appealing to users.

Conclusion

Despite the theoretical appeal that pricing bandwidth may have, it is very difficult to implement and account for. This difficulty is largely due to the underlying technology of the Internet. Packets are independent and the required bandwidth of a given user is bursty. Furthermore, both the sender and receiver of packets benefit, so it is difficult to determine who should be held accountable. Finally, it is difficult to monitor congestion on the Internet. Since TCP attempts to reroute packets once a given node is congested, points of congestion may be bursty.

In addition to the technical obstacles facing the service provider in pricing use, users may not want to take the time to figure out the advantages of a given scheme. In many analogous situations where the users are given the opportunity to choose among various forms of service, they opt for the simplest to understand. For example, even though it could be demonstrated to

---

[22] Ibid., 232.

users that they are better off under the expected capacity pricing scheme, they may still chose a flat rate. This phenomena is known as **Flat Rate Bias**. This bias may stem from another source. People simply don't like to be monitored while they are on-line.

Despite the economic ideal of implementing pricing to allocate bandwidth, the worlds of theory and practice often diverge. The only pricing scheme used to date is the flat fee. Access charges have no way of tracking the characteristics of transmission. As a result of the complexity associated with pricing schemes, there are several non-pricing schemes that may be used to curtail congestion. Although non-pricing schemes are easier to design, requiring less technical sophistication, they do come with their own unique problems.

B. Non-Pricing Schemes

1. Rationing The most obvious and common way to deal with scarcity is rationing. Rationing requires that bandwidth be assigned to certain users at certain times. Rationing may be applied in several contexts such as within networks, within organizations, or among service providers users.

Using rationing to allocate bandwidth is problematic for several reasons. First, it is likely to be perceived as being unfair and non-objective. Second, it is very difficult and perhaps costly for the rationeers to monitor and enforce. Finally, if bandwidth is not fully utilized during a given period as a result of rationing, when it otherwise would be, then rationing is inefficient.

2. Peer Pressure Another method of allocating bandwidth is to rely on peer pressure. Peer pressure requires that some users voluntarily forego use so that existing users do not experience depletability.

---

[23] Ibid., 236.

There are three problems with this approach. First, it only works when a very small number of users are involved, and the cost of planning among them is low. Second, as it becomes more impractical for users to plan, peer pressure may only be effective on an after-the-fact basis (opps. . . sorry. . . I was not aware you were also using the network). Third, there is no guarantee that users would cooperate. Again, the problem of enforceability is present.

3. Voluntary Control Mechanisms Voluntary control mechanisms require that users assign priority to their transmissions.[24] Consider an organization which is using a voluntary control mechanism. The organization may use three levels of priority for transmissions, urgent, necessary, casual. Upon sending a transmission, the user tags the transmission as being under one of the three categories. During periods of congestion, urgent transmissions receive highest priority at the router.

The problems of monitoring and enforceability are also present here. Specifically, a selfish user may assign all of their transmissions as urgent, regardless of content or purpose.

Conclusion

Although less technical to implement, non-pricing schemes do come with their own set of problems. Primarily, those of monitoring and enforceability. Even so, non-pricing schemes are used more than the pricing mechanism to alleviate congestion. Because of the problems of enforceability and monitoring associated with non-pricing schemes, use of them is primarily restricted to the organizational level.

II. THE LONG RUN PROBLEM

---

[24] Jeffrey K. MacKie-Mason and Hal R. Varian, 43.

Unlike the short run, there need not be a trade-off between depletability and excludability in the long run. In the long run, the Internet infrastructure no longer is fixed. Therefore, it is possible to provision the network to such an extent that congestion will not occur. As a result, there is no need to intervene to exclude users. In the long run, the Internet may be a true public good, possessing the qualities of nondepletability and nonexcludability. It is desirable to transform the Internet into a public good because its full value will be realized. As a public good, no performance degradation occurs and all users wishing to log-on may.

If everyone can use the Internet without performance degradation occurring, what's the long-run problem? The problem stems from the efficient pricing rule. The efficient pricing rule for a public good states that since there is no cost of supplying the good to additional users, it should be priced at zero. Note: in the previous statement, *cost* includes both the cost accruing to the service provider and the social cost of congestion. The long-run problem is finding a provider that will not bill for usage (not a very appealing proposition to a businessperson).

The difficulty in finding such a provider is compounded by the fact that backbone networks are extremely expensive to build. And because the associated costs are either fixed or sunk, they are independent of (efficient) pricing decisions.

A. Supplying Public Goods: A Historical Examination

The problem of finding a provider becomes clearer upon examining how public goods, and specifically the Internet, have traditionally been supplied. The national defense example is relevant here. Financing for national defense is supplied by the government and is paid for by tax revenues. The government also used tax revenues to fund the initial, and much of the subsequent, development of the Internet.

Financing a public good through the tax system may seem contradictory to the zero-price criteria. After all, citizens do pay for public goods, such as national defense and the Internet, with their tax money. However, tax based financing of public goods is not directly related to the intensity of use. The amount of taxes contributed by any given citizen is not associated with that citizen's use of the public good. That is, citizen A and citizen B may be in the same tax bracket; however, their level and intensity of use of the Internet may be very different.

A second method of financing public goods without violating the efficient pricing criteria is through fundraising. Fundraising is an informal type of taxation. Fundraising is voluntary, not compulsory. Furthermore, the amount contributed by any given user is determined subjectively, not through an objective formula, such as those upon which tax brackets are based.

If a public good is supplied through fundraising, some users of the service may never contribute. On the other hand, some users may contribute a disproportionately large sum relative to their use.

Oftentimes public goods are provided by combining formal taxes with fundraising. Examples are public radio and public television stations. Public stations do receive the minimum operating revenue requirement from the government. However, if they wish to provide higher quality programming, they may hold fundraising drives to bring in additional revenue.

B. Current and Future Concerns of Providing the Internet

Like many public policy discussions concerning resource allocation, determining how the Internet should continue to be supplied is a contentious issue. The debate is largely two sided. One argument is that the Internet should be developed under government supervision and financing. On the other side are the proponents of the markets ability to allocate resources efficiently. The question boils down to whether the private or public sector should be responsible for provisioning the Internet.

Both sides of the issue present strong arguments in support of their claim. The basic tenants of these arguments are not new to the public policy arena. Much of the argument in favor of the public sector provisioning the Internet goes as follows. If the government provisions the Internet, then they will do so in a fair manner. That is, the goal of provisioning would be to ensure that all members of society will have equal access to the Internet. Furthermore, all will experience the same performance quality.

Those in favor of private sector provisioning largely claim that the market can do it more efficiently and with better quality. This argument is valid on both fronts. For example, suppose the Internet were to be provisioned through a non-profit corporation. This corporation largely receives its funding from the donations from a diversity of private interests. This corporation is forced to be much more cost efficient than a government agency performing the same function. If the corporation is wasteful, its larger contributors will realize this and discontinue their contributions. However, a government entity that is wasteful may persist indefinitely.

With respect to the argument of technical superiority, only in isolated instances, such as national defense, is the government the technical leader. Even so, much of their technology is obtained from the private sector. Consider the non-profit corporation mentioned above. If this corporation operated for an extended period without adding to the body of technically expertise, again, its supporters would become fewer and fewer.

Another example is a large for-profit corporation provisioning the network. This company has a huge collateral base and has access to sophisticated methods of financing. Therefore, this corporation may be able to obtain the up-front costs of building a network.

Although they may not recoup revenues through use, they may have developed a creative way to bring in revenues associated with a different aspect of the Internet.

The argument against supplying the Internet through the public sector is that it may serve limited needs. It is not difficult to envision a scenario in which those with the deepest pockets and the most power receive the best service.

Conclusion

There is no clear black and white answer as to what sector should supply Internet capacity. Historically, it has been supplied by both the public and private sector. In the future such combinations are very likely. Such cooperation may in fact be the most beneficial method. Under a cooperative mission, the characteristics of each sector may serve to keep their respective evils in check.

Appendix

Case Study: Using the Internet for Voice Telephony

I. Traditional Model of Long Distance Access Charges

Traditionally, long-distance phone calls are charged for on a fixed fee basis, e.g., $.10/minute. However, the Internet may make it possible to forgo as much as 50% of this charge. Consider the scenario of a typical long-distance phone call made by a residential customer. Furthermore, assume the resident has separate local and long-distance carriers. The call proceeds as follows. The call leaves the resident's home over the dedicated phone line and travels to the local exchange carrier's (LEC) local office. The LEC then switches the call over to the long-distance (LD) carrier's network. The call then travels across to the LD network until it reaches the office of the LD carrier on the receiving end. The LD carrier on the receiving end places the call to the LEC on the receiving end, which in turn dials the resident on the receiving end to complete the call.

We know that the LD carrier is receiving $.10/minute from the resident for carrying the call. However, the LD carrier must pay the LEC for access to the LEC's facilities and services in placing the call. This charge is reasonable given that both the resident and the LD carrier are using the LEC's services to place the call. Given the $.10 LD charge, it is not uncommon for the LEC at the sending end of the call to require a $.03/minute access charge from the LD carrier. Furthermore, the LEC at the receiving end will require a similar charge. As a result, the LD carrier ends up paying the LECs $.06/minute or 60% of the revenue received from the resident.

This leaves the LD carrier with $.04/minute or 40% of per/minute long-distance revenue to pay for its cost of service.

II. Avoiding Access Charges: Theory and Practice

Theory The Federal Communications Commission (FCC) is the regulatory agency which regulates many aspects of the communications industry. Currently, the FCC has taken the stance that Internet service providers cannot be forced to pay access fees to LECs. As a result, there is an incentive for LD carriers to utilize packet-switched networks for voice telephony.

If a LD carrier were to build a packet-switched long-distance network to replace its long-distance circuit-switched network, it would be able to avoid paying access fees to the LEC. The procession of the call is very similar, but with a notable exception. Once the call arrives at the LD carrier's network, the information passes through a gateway that converts analog information into digital information. The message thus converted travels over a packet-switched network to the long-distance carrier on the receiving end. Once received, the information is converted back to analog.

By avoiding the LEC's access fees, the LD carrier could charge the resident \$.05/minute and enjoy \$.01/minute more than when using a circuit switched network. Although a \$.01/minute increase seems trivial, it does represent a 25% increase in per/minute revenue.

In addition to the incentive to avoid access fees, there are at least two other reasons to build packet-switched networks for voice telephony. First, it is technically more efficient to transmit information using packet switching. In fact, "a dedicated T1 can support as many as 130 IP (Internet provided) voice calls versus 24 simultaneous calls in today's carrier (circuit switched) networks."[25] That means packet switching technology can transmit over 500% more information over the same bandwidth as circuit-switched technology.

Second, packet-switched networks are cheaper to build than circuit-switched networks. This is because routers are currently cheaper than fiber-optic lines.[26] Circuit-switched networks use a lot of fiber-optic lines to "economize on switching and routing."[27] The benefit of circuit-switched networks are simplified accounting methods and routing calculations. "However, since about 1970, switches (computers) have become relatively cheaper than lines."[28] Therefore, it is more economical to place multiple routers performing switching calculations on a single fiber-optic cable. The disadvantage is that packets become difficult to account for and routing calculations are complex.

Practice In 1995, VocalTec Communications Ltd. introduced the first Internet-calling application.[29] This application enabled packet-switched voice communication between two personal computers (PC-to-PC calling). A person placing a call to another would have to speak into a microphone attached to a personal computer. The modem would then convert the analog message into a digital code so that it could be packet switched.

The technology of PC-to-PC calling has evolved into phone-to-phone Internet provided calling. By using the telephone rather than the modem, the technology has become more user friendly and reaches a broader segment of the population. In order to make an Internet voice call,

---

25 Dr. Matthew Lucas, "Billing Issues in International Internet Telephony," Billing World, November 1998, 34.

26 Ibid.

27 Jeffrey K. MacKie-Mason and Hal R. Varian, 36.

28 Ibid.

29 Lisa Bransten, "Net Effect," *Wall Street Journal*, 21 September 1998, sec. R23.

the user simply has to dial some extra digits at the beginning of the of the phone number. These extra digits are a code signifying that the transmission will be packet switched.

Although use of Internet provided voice telephony is on the rise, the quality of service is still an issue being addressed. From a technical standpoint, it is difficult to retain the qualities of analog transmission by providing voice communication over a digital network. Since packets are independent, do not arrive 'in order,' and may be delayed or dropped; the voice message being received may have poor and inconsistent sound quality in the form of delays, echoes and static. Despite this drawback, it is expected that sound quality will continue to improve as it has since its inception.

Not surprisingly, the ability to avoid access charges is an issue that LECs have taken contention with. LECs are providing the same service and incurring the same costs regardless of whether the call is circuit switched or packet switched. This is because the call is still circuit switched when it arrives at the LEC, and it is still circuit switched when it leaves. It can be certain that LECs will appeal to the FCC to have packet-switched networks placed upon a more equal footing with respect to access charges.

Perhaps as a result of such inequity, in an April of 1998 report, the FCC stated that it "reserved the right to look at data carriers and determine on a case-by-case basis whether they were providing information or telecommunications services."[30] Although the statement is general and does not make specific reference to access fees, it does provide for alternative policies in the future.

[30] Ibid.