

The history of communications and its implications for the Internet

Andrew Odlyzko
AT&T Labs - Research

amo@research.att.com

<http://www.research.att.com/~amo>

Preliminary version, June 16, 2000.

Abstract

The Internet is the latest in a long succession of communication technologies. The goal of this work is to draw lessons from the evolution of all these services. Little attention is paid to technology as such, since that has changed radically many times. Instead, the stress is on the steady growth in volume of communication, the evolution in the type of traffic sent, the qualitative change this growth produces in how people treat communication, and the evolution of pricing. The focus is on the user, and in particular on how quality and price differentiation have been used by service providers to influence consumer behavior, and how consumers have reacted.

There are repeating patterns in the histories of communication technologies, including ordinary mail, the telegraph, the telephone, and the Internet. In particular, the typical story for each service is that quality rises, prices decrease, and usage increases to produce increased total revenues. At the same time, prices become simpler.

The historical analogies of this paper suggest that the Internet will evolve in a similar way, towards simplicity. The schemes that aim to provide differentiated service levels and sophisticated pricing schemes are unlikely to be widely adopted.

Price and quality differentiation are valuable tools that can provide higher revenues and increase utilization efficiency of a network, and thus in general increase social welfare. Such measures, most noticeable in airline pricing, are spreading to many services and products, especially high-tech ones. However, it appears that as communication services become less expensive and are used more frequently, those arguments lose out to customers' desire for simplicity.

In practice, user preferences express themselves through willingness to pay more for simple pricing plans. In addition, there is a strong "threshold" effect to usage-sensitive billing. Even tiny charges based on utilization decrease usage substantially. In a rapidly growing market, it is in the service providers' interest to encourage usage, and that argues for simple, preferably flat rate, pricing. Historical evidence suggests that when service costs decrease, such arguments prevail over the need to operate a network at high utilization levels and to extract the highest possible revenues.

Communication services have long exhibited many characteristics of modern high-tech industries, namely high fixed costs, network effects, and difficulty in allocating costs among many products and services. Therefore the historical lessons of this paper are likely to be applicable to many pricing situations in the future. When prices are high, and purchases infrequent, airline-style "yield management" is likely to dominate. On the other hand, when prices are low, and a service is used many times a day, simple pricing and uniformly high quality are likely to be more common.

Historical analogies as well as current expenditure data also suggest that in the "digital convergence" of broadcasting and point-to-point communications, it is the latter that will dominate in shaping the evolution of the Internet. The current preoccupation with professionally produced "content" is probably more a distraction than help in planning for the future. Content has never been king, it is not king now, and most likely will not be king in the future. The development of the Internet is likely to be determined by the same growth of the myriad unpredictable commercial and social interactions that have fueled other communication services.

Contents

1 Introduction	3
2 Communications: grit and glamor, or why content is not king	12
3 Growth in communications: quantity and quality	31
4 The Internet and other communication services	45
5 The effectiveness and utility of pricing	56
6 The role of distance dependence in pricing	63
7 The conventional economic argument for flat rate pricing	67
8 The strong public preference for flat rate pricing	70
9 The dramatic effect of usage sensitive prices	72
10 Dynamic effects of flat rates	77
11 Miscellaneous examples of pricing trends	79
12 Mail	85
13 Telegraph	93
14 Wired voice phone	105
15 Cell phones	112
16 Minitel	116
17 Residential access to the Internet	118
18 Non-Internet data networks	120
19 Dedicated business connections to the Internet	121
20 Software agents and other technological palliatives	125
21 The role of distance distribution of traffic	127
22 The tragedy of the commons	131
23 Data transmission prices	133
24 Conclusions	137

The history of communications and its implications for the Internet

Andrew Odlyzko
AT&T Labs - Research
amo@research.att.com
<http://www.research.att.com/~amo>

1. Introduction

The goal of this work is to draw lessons from the history of communications. The view is broad, covering many technologies, and many decades or even centuries of data, and thus is of necessity not very detailed. However, it does produce some clear patterns. This is remarkable in that the history considered here covers many services, including regular mail, the telegraph, and the telephone. This history also covers many types of economic environments, including government monopoly, private untrammelled monopoly, government regulated monopoly, and competitive markets. The only constant factors have been relentless technical progress and growth in volume of communications. The historical record carries lessons for pricing, quality-differentiated services, the valuations of the dot-coms, as well as for the question of whether indeed “content is king” on the Internet. Many of these lessons carry over to general ecommerce.

The Internet has historically treated all packets equally, and pricing has been predominantly through flat monthly rates depending only on the size of access links, not on usage. Today, there is a strong momentum towards changing both of these principles. Influential figures in the data networking community make statements such as the following:

Although flat-rate continues to be the predominant form in which Internet access is sold, that form of pricing is unviable. Flat-rate pricing encourages waste and requires 20 percent of users who account for 80 percent of the traffic to be subsidized by other users and other forms of revenue. Furthermore, flat-rate pricing is incompatible with quality-differentiated services.

Pravin Varaiya, abstract of
INFOCOM'99 keynote lecture

There is extensive work on development of multiple service levels, typically with some implied guarantee of performance, usually referred to as Quality of Service (QoS). If those are deployed, they will necessarily require pricing schemes more complicated than the standard flat rate ones. The driving

force behind these developments is economic, based on the view that it is not feasible to provide satisfactory levels of quality for all transmissions at a reasonable cost. There is also concern (expressed in the quote above) about fairness of light users subsidizing heavy ones under the current scheme.

While the momentum is towards introduction of QoS and more complicated pricing systems, there have always been voices arguing in favor of preserving the simplicity of both operations and pricing of the current Internet. Their counterarguments can be substantiated through evaluations of the costs of multiple service levels to end users, and on considerations of what utility users derive from data networks [Odlyzko7, Odlyzko8]. Shortage of bandwidth is less of a problem than is often thought, and the general complexity of the system is a large barrier. An important observation behind many of the arguments in those papers is that what users really care about are fast transactions. As a result, whenever they can afford it, they run their links at low utilizations, since the main benefit they derive comes from the ability to burst at high speeds for short periods. In such an environment, flat rate pricing is most appropriate, and QoS does not help much.

The historical approach of this paper also leads to a skeptical view of most of the ongoing work on QoS. There have been many calls for usage-sensitive pricing similar to that of Pravin Varaiya quoted above. For example, an investigation of phone service in New York City in 1905 concluded

that, so far as large cities are concerned, unlimited service is unjust to small users, favors large users unduly, impedes expansion of the telephone business, tends to inefficient service, and that, as a financial proposition, is unsound.

p. 246 of [Stehman]

As will be shown in Section 14, the case for usage sensitive pricing in local phone calling in 1905 was much stronger than it is on the Internet today. However unsound it may have seemed, unlimited local calling for a flat monthly fee for residential users has persisted in most of the United States for all of the 20th century. It seems worth exploring why it did, what effects that has had in comparison with other countries that adopted usage-sensitive pricing, why it is being re-introduced in many of those countries, and what conclusions we can draw from this about the likely future of the Internet.

If we compare the telecommunications industries in different countries, we find few signs of harm from the “unsound” practice of flat rates for local calls. Table 1.1 shows that U.S. citizens use their phones considerably more than inhabitants of other affluent industrialized countries at only a slightly higher cost. Thus at least from this high level view, both consumers and service providers benefit from flat rate pricing.

Table 1.1. International comparison of telephone industry revenues and usage in 1997

country	revenues as fraction of GDP	minutes of phone calls per person per day
Finland	2.52%	16.6
France	1.93	10.6
Japan	2.06	10.6
Sweden	2.05	20.7
Switzerland	2.66	13.0
U.K.	2.51	12.7
U.S.	2.86	36.9

One can question whether history has much to teach us about the Internet. After all, incidents such as the transatlantic telegram of 1866 that almost bankrupted the U.S. State Department (see Section 13) show massive incompetence in cryptography and data compression as well as in basic business practices and diplomacy. Thus one can object that today we are much better prepared to handle a sophisticated technology. On the other hand, many of the practices that service differentiation would imply have been applied in the past. For example, Section 14 relates how over a century ago an Englishman was reprimanded for using his phone to summon the fire department to a blaze in a neighbor's house. History never repeats exactly, but there are frequent recurring themes that illuminate the present and hopefully can help make better choices for the future.

While QoS and sophisticated pricing schemes are widely regarded as necessary for the future of the Internet, the trend in other communication services has almost uniformly been in the opposite direction. Many examples are presented in later sections, and I will quote just a few here. In long distance voice telephony, the most popular plans are the simple ones that are independent of time of day or distance. Even in the Internet arena, the trend so far has been largely towards simple flat rates. A decade ago, the mass market online services, such as CompuServe, Prodigy, and AOL charged not just for minutes of connect time, but even for individual email messages. Email charges were eliminated first. By the middle 1990s, these services were also forced by subscriber complaints and competition from other ISPs to switch to unlimited access for a flat monthly fee.

Although simplicity in pricing has been the general trend, it is important to recognize that there is persuasive logic to differentiated services. Even in the absence of quality differentiations, there are good reasons for pricing according to usage, as is done with electricity. Some of the arguments in this direction will be reviewed later. A major problem with such measures, though, is that usage sensitive

charging, even with very low rates, lowers usage, often dramatically. Sometimes that is desirable. Sometimes, though, it goes against public policy or against interests of service providers. Western Europe is beginning to re-introduce forms of flat rate local phone calling precisely in order to stimulate Internet usage, which lags behind that in North America. Numerous examples are presented later to show that the best way to stimulate usage is to have flat rates. The Internet is expected to continue growing and to penetrate ever more deeply into our lives. Hence the competition among ISPs will be won by those who can entice their subscribers to make the fullest use of the services that are offered. Flat rates or approximations to them are among the most powerful tools in this competition.

The general trend has been for each communication service to offer higher quality and lower prices. This has gone hand in hand with higher usage, and higher total revenues. (The telegraph is a partial exception to this trend, and it is instructive to study its history for that reason.) At the same time, pricing has tended to become simpler.

A particularly telling instance of the trend towards simplification in communication is the tendency for prices to become distance-insensitive as a service evolves. The basic justification for the development of quality-differentiated service levels on the Internet is the perception that transmission and switching resources are expensive, and therefore high quality cannot be provided uniformly for all traffic. Yet that does not seem to have been a barrier for other communication services. The distance dependence in pricing of mail, telegraph, telephone, and even data network services has historically decreased, or even disappeared completely. A Martian who saw just the evolution of price schedules might be tempted to argue that if the extra cost of network facilities to send messages to distant locations is not worth charging for, then the extra cost of overprovisioning the network to provide uniformly high quality to all transmissions is likely to be negligible as well.

As will be explained in Section 6, distance dependence in pricing is often a sign not of substantially greater costs for long distance service, but a form of price discrimination. Long distance transmission is regarded as more valuable. Thus the decrease in distance dependence of pricing provides an even stronger argument than that of the hypothetical Martian. Perhaps an even more interesting point of view goes even deeper into an analysis of possible forms of pricing. A schedule of charges based simply on distance is already a major concession to the desire for simplicity, and away from optimality in either revenue extraction or in relating prices to costs.

One can take the historical evidence of this paper that communication pricing has been evolving towards simplicity and argue that by simple analogy, the Internet is also likely to move in that direction. A sounder procedure is to investigate more carefully what factors contributed to historical developments,

and what influence they might have in the future. There are, after all, reasonable arguments against simplicity. A large part of this paper is devoted to an investigation of the reasons for various pricing strategies.

Sophisticated strategies for quality differentiation and price discrimination are common. They are perhaps most noticeable in airline yield management, but are spreading to other areas. For example, Coca Cola is experimenting with vending machines that will automatically raise prices when temperatures are high. We can expect such practices to be widely adopted for two main reasons. One is that the economic evolution of our society is increasing the role of fixed costs in the provision of goods and services. Therefore pricing on the basis of marginal costs is becoming untenable, and it becomes necessary to price on the basis of customers' willingness to pay. That calls for approaches such as those of airlines and Coca Cola. The other reason we can expect such practices to spread is that modern information technology is making them possible. In the past, Coca Cola might have wanted to price its drinks depending on customers' thirst, but could neither predict the degree of that thirst, nor could it adjust prices in a timely fashion. Now it can do both.

While there are sound arguments for quality differentiation and price discrimination (treated briefly in Section 5), there are also economic arguments against them (described in Section 7), based on economic models that favor bundling. Perhaps more important, such practices conflict with the strong preference that people have for simple pricing, especially for flat rate schemes. (Section 8 is devoted to this subject.) There are many reasons for this preference. The insurance aspect is important ("How do I know how high a bill my teenagers will run up?"). So is another aspect, neatly summarized by Nick Szabo's phrase [Szabo], "mental transaction costs;" "Yes, I can save by optimizing my usage, but do I want to, if the savings come to fractions of a penny, and require my attention dozens of times a day?"

The main point of this paper is to observe how the conflict between the need to optimize and people's reluctance to optimize has been resolved historically. The general conclusion is that complicated pricing and quality differentiation strategies have been applied and continue to be applied for expensive or infrequent purchases. On the other hand, the search for optimal efficiency has usually been abandoned in favor of simplicity for small and frequent transactions. The long history covered in this paper illustrates how the decreasing prices of communication services, the fruit of improving technology and of economies of scale, have gone hand in hand with the trend towards greater simplicity.

It seems reasonable to expect that a similar division between the quest for optimal efficiency on one hand, and simplicity on the other, will hold in the future. It will be easier for service providers to implement complicated pricing schemes. There will also be more tools (and intermediaries) helping

consumers to navigate the maze of choices. However, both factors have been operating for a long time and that appears only to have strengthened people's desire for simplicity. The Internet is leading to an unprecedented proliferation of choices in goods and services, straining human capacity to deal with all the complexities. (As an example, voice response systems are used widely, yet have not diminished the need for human assistance. Instead of just simplifying old tasks, they have led to more complicated choices, which require human assistance on top of the automated systems.) Therefore the balance between optimization and simplicity is likely to remain where it has been, in spite of the developments in technology. That is a major reason that airline tickets are much more likely to be sold through complicated schemes than voice calls. Even in the U.S., which is affluent, and has a large air transport industry, there is only about one round trip per person per year by air, at a cost of around \$400. Thus airplane tickets are expensive and are bought relatively infrequently. Further, air transport is a mature industry, and costs are declining slowly. On the other hand, on average each U.S. inhabitant is engaged in about a dozen phone calls per day, each typically lasting about three minutes, and costing either nothing or a fraction of a dollar. How much thinking about the price and value of each call would people want to do? The same arguments apply even more strongly to Internet usage. Will we want to be bothered about the cost of each Web page download, or any of the myriad other small transactions we will be engaging in? Moore's Law is providing ever powerful silicon chips, but the nerve cells in our brains are not getting any faster. Human attention may be the ultimate scarce resource.

The historical trend of increasing volumes and decreasing costs in communications is illustrated by comparing the purchase of a book today to the receipt of a letter in England in the the early 19th century. In 1834, about 5 letters were mailed per person per year in England. In most parts of the country, the recipient had to learn that a letter was waiting, go to the post office and pay a fee that averaged what about two hours of work cost. (See Section 12 for more details.) In contrast, in the U.S. of 1994, there were about two pieces of mail delivered per person per day. The cost of each piece was under two minutes' wages for an average worker. That same year, 1994, the U.S. book industry sold 2.13 billion books, or about 8 per person. The average price was \$11.19 per volume, a little less than the average hourly wage. The full cost of the acquisition also included going to the bookstore to find the book, pay for it, and so on. Today, half a dozen years later, we are in the in the Amazon.com era. Those additional non-monetary book purchase costs can be decreased, as we can order books with a few keystrokes and mouse clicks, and have them delivered to our house a day or two later. Thus in terms of relative costs, books today are acquired more often and cost less than letters 170 years ago.

As costs decrease, and volumes increase, it is only natural that the care we devote to each message

or transaction goes down. Thus it is to be expected that we should be less willing to spend time and energy optimizing our transactions. Section 3 reviews the remarkable and continuing increase in the volume of communications, and the effects this has had on the care we put into individual messages.

Before considering growth and price statistics, in Section 2 we look at the relative values of different types of communication. Currently there are widespread expectations that “content,” usually denoting professionally produced multimedia traffic, will dominate the Internet. However, even if the volume of multimedia traffic does exceed that of other transmissions, designing the network primarily for such traffic is likely to be a mistake. Historical evidence, including current statistics, shows that people are willing to spend far more on point-to-point communication than on entertainment. Thus it is unlikely that professionally produced multimedia traffic will determine the evolution of the Internet. To continue attracting the resources that we are devoting to communications, the Internet will have to cater to user desires for flexibility, universal reach, informal associations, and other spontaneous interactions that no “content” can satisfy.

If the above prediction about secondary role of content comes to pass, it will fit a historical pattern. For example, in the early 19th century, U.S. postal policy was preoccupied with broadcast-type communication, namely through newspapers. Distribution of newspapers was subsidized to an extraordinary degree by regular letters, which brought in most of the money. Eventually public pressure forced policy makers to lower the prices of letters, which led to an upsurge of traffic.

The next few sections review some of the main arguments for and against price and service quality differentiation. Section 11 then discusses pricing trends in several areas related to communications. Finally, Section 12 starts consideration of communications by examining the history of the postal system. This is continued in the following sections on the telegraph, the telephone, and various related services.

Historical analogies are interesting, but may not always apply to a new technologies. The Internet is unique, and some of the important ways it differs from other services are examined in sections 20-23. Perhaps most important, prices have been increasing in recent years for most of the Internet. This is a seldom-noted phenomenon that is almost unique among modern high-tech industries in which demand has been increasing and technological progress has been stunningly fast. It is likely that this increase in transmission prices has been responsible for much of the demand for QoS and complicated pricing. Once prices start plummeting overall, as they have begun to do on some routes, it seems likely that providing more bandwidth will be seen to be more attractive than forcing complicated schemes on users.

There are two other ways in which the Internet is often claimed to differ from other communication

services. One is in the possibility for a “tragedy of the commons,” in which flat rate pricing leads users to saturate any links that are offered. The other is in the independence of traffic from distance. However, as sections 21 and 22 show, neither of these is likely to be a serious problem for the Internet.

The general conclusion that I draw from the evidence presented in this paper is that the Internet is not unique. Most likely it will continue to evolve the way other communication services did, namely towards higher quality for all transmissions, lower prices, and simple prices. We are likely to end up with a mix of traditional flat rate offers, together with some form of the “expected usage pricing” scheme proposed in [Odlyzko8]. That proposal is discussed in greater detail in Section 5. In this model, a customer (residential or business) would be offered unlimited usage for a year, say, with the price determined just by the capacity of the link and that user’s record. Service providers would act as insurance companies, using prior usage, demographics, and other information to fashion customized bundled offers that would meet the demands for simplicity and flexibility.

This paper pictures pricing of communication services as the outcome of a competition between the need to optimize on one hand, and to cater to user preferences for simplicity on the other. This view is unconventional. Most standard works, such as [MitchellV], concentrate on optimizing some conventional quantitative utility function, and treat customer desires as irrelevant, or else as irrational annoyances that interfere with clever and efficient schemes. In particular, such works treat preferences for flat rates dismissively, if at all. For example, we find in a paper on pricing of local phone service the following statement:

... Clearly a movement to a positive per call charge would increase aggregate economic efficiency. Yet nearly all proposals for a move to [usage-sensitive pricing] have met stiff consumer resistance. The reluctance seems to persist even when customers face the prospect of a [usage-sensitive pricing] plan that would, on average, result in a lower monthly bill.

[Panzar]

That paper then goes on to propose a usage-sensitive pricing plan that would hopefully help wean customers from their apparently irrational reluctance to embrace such schemes.

This paper approaches the problem differently. It presents extensive historical and current data about customer preferences, explores the reasons for such preferences, and considers the effects of catering to them. It argues that however irrational such preferences may seem, they should be paid attention to. It shows that, generally, such a course of action is likely to be best for rapidly growing communication services like the Internet. Technology is boosting the capacity of all networks. There-

fore customers have to be encouraged to increase usage, to fill this growing capacity. Carriers that fail to persuade their customers to move on to higher bandwidth services will be left behind in the competitive race. Flat rate pricing is optimal in encouraging use. For cases where flat rate pricing plans are not feasible, this paper also suggests usage sensitive plans that try to appear as close to flat rate to the user as possible.

Standard works on telecommunications economics usually say nothing about the effect of flat rates on usage. The only work I have found that treats this subject in any detail is the rather obscure report [Kraepelien].

This is obviously not the definitive history of communications. No work could possibly fill that role. Even a complete coverage of communication pricing, the subject of most of this paper, is impossible in a work of this scope. (The book [Smith] devotes over 400 pages to a study of postage rates alone, and yet covers only a few countries and only up to World War I.) In particular, practically no attention is devoted to technology, regulation, industrial organization, and many other crucial aspects of communication services. Their effects are noted only through the decreasing prices and increasing quality and quantity that users experience. Instead, I pick out a few threads from this rich subject that combine to present a coherent picture of relevance for the Internet. The main themes are growth of communication, the qualitative change this growth produces in how people treat communication, and the evolution of pricing. The stress is on the user, and in particular on how quality and price differentiation have been used by service providers to influence consumer behavior.

The selection of historical data in this paper is heavily weighted towards the early periods of each technology that is considered. One reason is that those early periods tend to be more colorful. For example, today telegraph company lawyers do not hire people to cut down the poles of a competitor (as happened in England in 1861, see p. 63 of [Kieve]). Also, railroads that compete for the same bottleneck mountain pass do not hire private armies that engage in shooting wars (as happened in the infamous 1878-1879 contest between the Denver and Rio Grande Railroad and the Atchison, Topeka, & Santa Fe). A more substantial, but related, reason for concentrating on early stages of communication services is that they help clarify the evolution of those services. The contrasts between the early and mature stages of a technology bring into sharper contrasts the issues of how people treated those services, and what the role of pricing was.

The general conclusion is that the forces that have worked to produce simplicity in pricing in the past are likely to be at least as influential on the Internet. Optimal efficiency will be sacrificed in the interests of satisfying customers. Further, professionally produced “content” will not be the main

factor determining the evolution of the Internet. Communications is likely to grow in revenues and importance, but will do so in often unpredictable ways, as society adopts the new tools to suit its needs.

Note: A section at the end, between the Acknowledgements and the References, lists sources for the data presented in the tables. It is worth emphasizing now that all the monetary figures are in current dollars, and thus are not adjusted for inflation. Price indices for the 19th century are rather imprecise, and in any case inflation-adjusted costs do not present an accurate picture of the burden of prices back when living standards were much lower. Therefore some prices are stated in terms of hours of work required to pay that price.

2. Communications: grit and glamor, or why content is not king

The Internet is widely predicted to produce “digital convergence,” in which computing, communications, and broadcasting all merge into a single stream of discrete bits carried on the same ubiquitous network. The popular images of convergence are heavily tinged with the flavor of Hollywood. “Content is king” is the universal buzzword, where content is usually taken to mean professionally prepared material such as books, movies, sports events, or music. The race is supposedly to determine which organization or alliance will dominate in providing content to the users, ideally in advanced multimedia formats. A recent article concludes that “[the Internet] has become a mass medium used mostly by relatively passive consumers, and as such major content providers will dominate it” [MargolisR]. The book [Winston] also presents the Internet as the next step in the evolution of mass media. Many industry leaders appear to base their strategies on this thesis. For example, at Global Crossing, its latest CEO, Leo Hindery, is attempting

to turn this global Internet-based network into a mature content distributor. ... “I don’t want to be anyone’s dumb pipes,” says Hindery. “If all you do is racks and servers, that’s dumb. What we’re doing is melding the network and the content.”

[Krause]

This preoccupation with content is not peculiar to North America. Norio Ohga, once CEO and recently chairman of Sony, says that “[w]ithout content, the network is nothing” [Schlender]. Juan Villalonga, the chairman of the dominant Spanish communications carrier Telefónica, bases his strategy on the belief that “[t]he key ... is content. Without it, ... phone companies risk becoming simple commodity pipelines” [Baker].

Unfortunately for these companies, content is not the key. Content certainly has all the glamor. What content does not have is money. This might seem absurd. After all, the media trumpet the hundred million dollar opening weekends of blockbuster movies, and leading actors such as Julia Roberts or Jim Carrey earn \$20 million (plus a share of the gross) per film. That is true, and it is definitely possible to become rich and famous in Hollywood. Yet the revenues and profits from movies pale next to those for providing the much denigrated “pipes.” The annual movie theater ticket sales in the U.S. are well under \$10 billion. The telephone industry collects that much money every two weeks! Those “commodity pipelines” attract much more spending than the glamorous “content.”

A reasonable objection to the comparison above is that what matters in directing investments is profits, not revenues. That is true. However, one cannot have profits without revenues. Further, profits of the telephone industry have dwarfed those of Hollywood. Even if we look at profitability in terms of return on investment, it is not clear that movies have been notably more profitable than communications, especially if one adjusts for risks. Those who invested in Disney in the early 1980s (but not recently) have done very well, but Sony took a bath in its takeover of Columbia Pictures. Some creative talent has done very well. An outstanding example is Steven Spielberg, who became a billionaire, while minimizing his risks through careful structuring of the deals. On the other hand, for most actors and writers, the financial rewards are much slimmer and spottier, as Hollywood is very much a “winner-take-all” market. In communications the risks may also be rising, as the Iridium debacle demonstrates, but so are potential returns. At this moment Wall Street gets attracted primarily to the prospects for rapid growth, figuring that profits will show up some time in the future. While the actual returns that Wall Street seems to be expecting may be ludicrously overoptimistic, the general principle appears valid. The histories of the telegraph and the telephone show the same pattern. Usually many companies jumped in with unrealistic hopes, most failed, but the industry as a whole prospered. Therefore the rest of this section will concentrate on revenues and growth rates of various sectors of the high-tech economy, and what they say about the current and potential role of content.

Table 2.1 presents statistics that show the relative sizes of several sectors of the U.S. economy. The data was drawn primarily from [USDOC2], and the year 1997 was the most recent for which all the relevant time series were available. The precise description for how the figures were obtained is given in the section just before the References. There is considerable overlap in different categories in Table 2.1. For example, the \$187.5 billion of advertising industry revenues pays for almost all television broadcasting, and that provides much of the funding for the movie industry. Further, consumer expenditures on phone services are already contained completely in the general telephone industry figure.

Some of the categories, such as sporting goods and airlines, are included just for illustration.

Table 2.1. Selected sectors of U.S. economy.

industry	1994 revenues (billions)	1997 revenues (billions)	annual growth rate
telephone	\$199.3	\$256.1	8.7%
long distance	81.0	98.5	6.7
wireless	16.8	33.5	25.9
U.S. Postal Service	49.6	58.3	5.5
advertising	151.7	187.5	7.3
motion pictures	53.5	63.0	5.6
movie theaters	6.2	7.6	7.0
video tape rentals	7.0	7.2	0.9
broadcast industries			
television broadcasting	31.1	36.9	5.9
radio broadcasting	10.5	13.5	8.7
newspapers	47.2	55.3	5.4
magazines	17.4	19.9	4.6
consumer spending on "content"	113.9	133.5	5.4
subscription video	29.2	41.5	12.4
home video (rental and purchase)	17.8	20.4	4.6
home video games	3.1	4.4	12.4
newspapers	12.8	13.6	2.0
consumer magazines	9.5	10.1	2.1
consumer books	20.2	20.9	1.1
recorded music	14.7	14.9	0.5
consumer spending on phone service	70.5	85.4	6.6
sporting goods sales	53.5	64.1	6.2
airlines	88.3	109.5	7.4
national defense	281.6	270.5	-1.1

What is striking is how highly valued communications is. If we combine the revenues of the phone industry with those of the postal service, we obtain a figure larger than military spending, and almost

three times higher than the revenues of the airline industry. Just the spending on phone services is higher than all advertising outlays. So say good-bye to all those plans for financing the Internet through advertising! Yes, advertising can help fund some services, but it will not provide the generous revenue streams that are needed to support a communications infrastructure as large as the phone system. To obtain the funding that many dot-coms seem to be planning on, it will be necessary to get contributions from more than advertising. Ecommerce can help, but even that probably will not be enough, and it will be necessary to persuade people to pay for a large chunk of their communications. The question is, what are people willing to pay for?

Table 2.1 shows that some advertising-supported business models might indeed be feasible. For example, sales of recorded music come to about \$15 billion per year. If one eliminates the overhead costs of the physical distribution system for CDs, one could probably provide the artists with as much money as they make now, and the music labels with as much revenue for their central selection and promotional activities (and their profit) for under half of the \$15 billion. That would be about half of the advertising revenues that the radio industry collects for broadcasting music. Getting that much extra funding from advertising might be possible for an Internet music service that allowed listeners much greater selectivity and thereby led to more listening (as Napster appears to be doing on college campuses). However, such a move appears feasible only because recorded music is a relatively small market. We could not hope to obtain enough advertising money to pay for anything as large as the phone system.

Although Table 2.1 makes a powerful case by itself, it is worth reiterating the basic theme, which is that the vaunted “content” is not where the action is. The postal system alone collects almost as much money as the entire movie industry, even though the latter benefits from large foreign sales. For all the publicity it attracts, entertainment is simply not all that large, because people are not willing to pay very much for it. The dream of the early 1990s of financing the “Information Superhighway” through “500 channels to the home on the cable TV network” was an obvious fantasy.

Content is not only a small part of the economy, it is often paid for indirectly. Well over two thirds of newspaper revenues, and almost all of broadcast TV and radio revenues come from advertising. Thus content is being given away in order to attract people to goods and services they are willing to pay for. Although spending for content (whether by consumers or advertisers) has been rising, it has been doing so at a sedate pace, and is unlikely to explode.

One could object that Table 2.1 proves just the opposite of what is claimed above. After all, this table shows that even if no single content segment collects anywhere near as much money as the phone

system, in aggregate huge sums of money are being spent on content. In particular, consumer spending on content is over 50% higher than on phone services. There is some issue of what one means by content, a question we will return to later. If we take a generous interpretation, we can come up with total content industry revenues comparable to the \$256 billion that the telephone industry collected in 1997. (This would include consumer spending as well as business information services and advertising revenues of broadcast industries.) However, comparing just the total revenues of those two industries is misleading. In the case of the telephone industry, the \$256 billion does include some service revenues as well as yellow pages advertising, but the overwhelming majority of that money is for simple transport of voice and data. The content industry as a whole, though, has to use its revenues to pay for content as well as the delivery of the content. The creative souls who command \$20 million per movie know their value, and do manage to appropriate a lion's share of the profits from such enterprises, while keeping their risks lower than those of the investors. Similarly, the professionals who compile the Lexis database, or assemble and monitor Reuters' financial data feeds, have to be paid, just like movie actors and musicians. Even if their average pay is lower, there are more of them, and their payroll, as well as all the equipment and overhead needed to support their work, are not inexpensive. Hence only some of the revenues of the content industry contribute to the communications infrastructure. Since this work is concerned with the future of the Internet, it is the present and potential funding for the network that matters.

That only a fraction of the revenues of the content industry go for delivery is an important point that has analogs in the context of ecommerce. Ecommerce is already big, and is exploding. However, what does that mean for the network? A dollar of ecommerce transactions does not mean a dollar devoted to the network. When that sterling example of ecommerce, Amazon.com, sold \$1.6 billion of goods (primarily books) in 1999, it is likely that only a few million dollars of the \$2.3 billion of its costs went for Internet connectivity. Considerably more, but still only around \$150 million, went for the servers, software development, and other information technology products and services that are needed to stay competitive in this rapidly changing field. If the Amazon figure of \$1.6 billion is to represent the ecommerce opportunity for the Internet, then t-commerce ("t" standing for "telephone", and covering all deals that use the phone in any way) amounts to tens of trillions of dollars. (Yes, more than the GDP, since the wonderful accounting of the ecommerce world surely would let us count the same value several times, as it passes through multiple transactions that all use the telephone at some stage.) The contribution of ecommerce to communications is growing, but it has to be kept in perspective.

Communications is far from being the largest segment of the economy. It is smaller than cars, housing, food, and especially medicine. It is also about two thirds the size of the primary and secondary education sectors, and comparable to the higher education enterprise. The main point, though, is that communications is huge, and represents the collective decisions of millions of people about what they want. It is also growing relative to the rest of the economy in a process that goes back centuries, as the next section will show. As a fraction of the U.S. economy, it has grown more than 15-fold over the last 150 years. The key point of this section is that most of this spending is on point-to-point communications, not for broadcast media that distribute “content.”

The predominance of point-to-point communications spending is not new. That has been the historical pattern for ages. For example, in the early 19th century, almost all the revenues of the U.S. postal system came from letters. Yet about as many newspapers as letters were being delivered.

The preoccupation of decision makers with content and broadcast communication is also not new. In the early 19th century, the explicit policy of the U.S. government was to promote wide dissemination of newspapers. They were regarded as the main tool for keeping citizenry informed and engaged in building a unified nation. Hence newspaper distribution was subsidized from profits on letters, as will be discussed at greater length in Section 12. The extent of the subsidy may be gauged by the fact that “[i]n 1832, newspapers generated no more than 15 percent of total postal revenues, while making up as much as 95 percent of the weight” (p. 38 of [John1]).

The policy of the U.S. government to promote newspaper “content” at the expense of person-to-person communication through letters may or may not have been correct. It would be a hard task (and one well beyond the scope of this work) to decide this question. However, there are reasonable arguments that the preoccupation with newspapers harmed the social and commercial development of the country by stifling circulation of the informal, non-content information that people cared about. In the 1840s, responding to public pressure, Congress did reduce letter rates, which resulted in increased usage, and changed patterns of usage, as is described at greater length in Section 12. In those days, the government understood clearly that what people were willing to pay for was letters, and that newspapers were being subsidized. The Post Office would have thrived on letters alone, but would have gone bankrupt instantly had it been forced to survive on newspaper deliveries. Thus content was king in the minds of policy makers, but it was definitely not king in terms of what people were willing to pay for. That is similar to the current situation. However, this differential in willingness to pay does not seem to be understood as well today as it was then.

Preoccupation with content has historically been common. For example, it was often thought (even by Alexander Graham Bell) that one of the principal uses of the telephone would be in broadcasting [deSolaP1, deSolaP2]. Several substantial experiments in delivering content over the phone were attempted, including the Telefon Hirmondó in Budapest that lasted from 1893 past the end of World War I, and the Telephone Herald in Newark, New Jersey, which folded soon after its start in 1911 [deSolaP1]. In the end, though, the phone emerged as the prototypical example of point-to-point communication.

Radio initially evolved in the other direction. It started out as a point-to-point communication technology, a wireless telegraph. After about two decades of experimentation, it became primarily a broadcast medium. (For the history of this transformation, see [Douglas, Smulyan].) However, the role of radio in the economy as a content delivery technology is tiny compared to that of the telephone, as Table 2.1 shows. Further, in the last few decades, with the development of cellular services, radio transmission has started to move back to its roots as a point-to-point communications service. The revenues from wireless telephony already far exceed those from radio broadcasting, as Table 2.1 shows (\$33.5 billion versus \$13.5 billion in 1997, with the disparity much greater today).

A skeptical reader might say that all this historical stuff is amusing but irrelevant. We live in the 21st century, and our high-tech present as well as our future are on the Web, where content is universally regarded as king. Studies of the Internet regularly find that Web traffic makes up 60 to 80% of the bytes that are transmitted. Certainly most of the commercial development effort on the Internet and almost all the attention are devoted to content. Thus even if content was not king in the early 19th or late 20th centuries, it might be king in the 21st.

There are three counterarguments to the above objection, all of which support the “content is *not* king” thesis. All argue that the dazzling success of the Web has created a misleading picture of what the Internet is, or is likely to evolve towards. One argument, to be discussed in more detail later, is that the future of the Internet is not with the Web, but with programs like Napster or (even more, because of its decentralized nature) Gnutella, which allow for informal sharing of data.

The second argument is that content is not king of the Web. Most of the traffic on the Internet is corporate (especially if we include internal intranet traffic that is not visible on the public backbones). It is likely that in early 2000, no more than a quarter of the volume went to residential users [CoffmanO, CoffmanO2]. Intranet traffic appears to be much less heavily biased towards the Web than that of private individuals. Furthermore, even the traffic that appears to be Web-based frequently represents a variety of database transactions that are not properly speaking “content.” Because browsers are a

user-friendly tool that is ubiquitous, a multitude of services have been squeezed into a Web framework, and help perpetuate the image of the Internet as primarily a content-delivery mechanism.

The third and final argument is that even if content were king on the Web now, the Web is *not* king of the Internet. This may again seem absurd, especially in view of the statistics quoted above, that most of the Internet traffic is Web transfers. However, consider again the U.S. postal system of 1832. Content certainly dominated in terms of volume of data. Newspapers sent by mail weighed about 20 times as much as letters. Further, the density of printed matter is higher than of handwriting, and a typical copy of a newspaper was likely read many more times than a typical letter. Hence newspaper “content” was probably delivering at least a hundred times as much information as letters. But volume is not the same as value. Letters were bringing in 85% of the money needed to run the postal system in 1832. On the Internet in 2000, it is email that is king, even if its volume is small.

Today, Web traffic dominates the Internet in volume, with about 20 times as many bytes as email. (Netnews traffic volume is of about the same order of magnitude as email. The fractions of total traffic created by these two services vary from link to link and from day to day.) Even a decade ago, before the Web, email typically accounted for under 10% of Internet volume. Yet email has been and continues to be the real “killer app” of the Internet. The ARPANET (the progenitor of the Internet) was built primarily to connect computers. Yet email quickly emerged as the application that mattered the most to users, even in the early days of the network. This was much to the surprise of the system’s designers [LickliderV]. More recent surveys (e.g. [KatzA]) show that email is still the most valuable service. Ask people whether they would rather give up email or the phone, and the responses will typically be split. However, when a similar choice is offered between the Web and email, there is no contest. This is true for both individuals and large organizations. Intranets are all the rage, but it is email that makes enterprises run.

The perception that content dominates the Internet is fueled by studies such as [AdamicH], which show the winner-take-all phenomenon, with a few sites dominating Web transactions among residential users. However, one should not read too much into such results. In the early 19th century postal system, studies of usage of information would undoubtedly have reached conclusions similar to those of [AdamicH]. Most data from distant locations that people consumed came from newspapers. Further, circulation figures of individual newspapers probably followed the standard Zipf-type distribution, with the most popular papers attracting a disproportionately high fraction of readers. Yet that does not say much about the value derived from the postal system, which was elsewhere, in letters.

To reemphasize the importance of point-to-point communication in the online environment, consider the disappointing reception that WebTV has had. It seems that inexpensive Web browsing is not such a great attraction by itself. Also consider the innumerable failures in teletext experiments (cf. [Ettema, Greenberg, Klopfenstein]), as well as the initially promising start but disappointing end of the French Minitel (dealt with in more detail in Section 16). Their inadequate or even totally missing facilities for point-to-point communication appear to have been fatal errors. Next, consider the fates of CompuServe and Prodigy. Set up primarily for database access and online shopping, respectively, both were forced to emphasize communications and are now basically standard ISPs. Finally, the currently most successful of the public online services, AOL, started out as a game network. The figure in Table 2.1 showing small video game spending explains why that approach was doomed to failure. (The game market is growing, and in any case is larger than the figure in Table 2.1, which covers just the home part of it. It can certainly contribute to profits of large networks, or support some specialized service providers. However, it simply cannot fund anything as large as AOL.) AOL survived and prospered because it was nimble. After several changes in strategy, it partially opened itself up to the Internet. While it has content of its own, and access to the Internet, the majority of the time its subscribers spend online is devoted to email and chat.

What this argument suggests is that the Web (and browsers in particular, which made the Web user-friendly) may have created a misleading impression. By focusing attention on centralized delivery of content, the Web may have prevented a proper appreciation of the importance of the often chaotic and generally unplannable point-to-point communications. The Web and the browsers may have played two main roles. One was to force online service providers to accept an open interoperable standard that made the entire Internet accessible for communications for everyone. The other was to introduce a user-friendly graphical interface for email, chat, and netnews, which made such communications easier. However, the Web is probably not as important to the Internet as is commonly thought.

As an example of the relative value of content and simple pipes, note that the revenues of the entire cable TV industry in 1997 were only slightly higher than for the cell phone carriers. Furthermore, as Table 2.2 shows, cable TV has been growing far more slowly than the wireless industry. (See Table 3.5 for more detailed historical statistics for wireless communication.) Yet cell phones currently provide primarily simple, low bandwidth pipes. By early 2000, even the gross revenues of radio telephony exceeded those of the cable TV industry. Further, a large chunk (estimated at a quarter to a third) of cable revenues was devoted to paying for content, so in terms of basic network revenues, the cellular industry had pulled ahead even before 1997. The comparison is even more favorable to low bandwidth

wireless pipes when we go outside the U.S. In other industrialized countries, cell phones are often much more widely used, while cable TV penetration is almost always far lower than in the U.S. Thus on a worldwide scale, the comparison is skewed even more heavily against content. The cable TV industry does have excellent prospects for faster growth. However, that growth will surely come more from improved communication capabilities, and less from content.

Table 2.2. Revenues of U.S. cable TV and cell phone industries.

year	cable TV (millions)	cellular (millions)
1987	\$11,563	\$942
1992	21,079	6,688
1997	30,784	25,575

The general conclusion is that content has been less important than point-to-point communication in the past, including the recent past involving the Internet. Still, the argument that “content is *not* king” that is presented here should not be taken to an extreme. All it says is that most of the money is in point-to-point communication. It does not say that content does not dominate in volume of data. Historically, as we have noted above, content has often dominated, and probably dominate now. (There is some uncertainty, since there are difficult questions about measuring the volumes of broadcast communications.) There are arguments, to be presented later, that in the future, content will not provide most of the bits traveling on the Internet. However, even if that prediction is wrong, it will not affect the argument, which is about value to customers, and not about volume.

That content is not king does not mean that content is unimportant in shaping political or social views. The attention paid to writers, and the political advertising on radio and TV, testify to the influence of content. This also is not a new phenomenon. Over the centuries, millions of people based their opinions of Richard III on Shakespeare’s play, just as today millions base their opinions of John F. Kennedy’s assassination on Oliver Stone’s film.

The argument about the value of content says little about the dispute in early 2000 between Disney and Time Warner over carrying ABC channels on cable TV. That issue is about division of revenues between content creation and content distribution. The argument of this paper is that the entire content piece of the economy is not all that large, and its contribution to network costs is much smaller than that of point-to-point communication. It does not deal with how the content piece is divided.

Content can be profitable. Numerous media companies are doing very well. Content can also be of value to a network, even aside from providing traffic for the network to carry. However, it is probably best to think of content as either catnip or icing on the cake; something to attract new users, or enhance user experience. That is what broadcast TV programs do for the advertisers who pay for them. That may also have been the main role of the Web and browsers in bringing more people to the Internet.

What the argument that content is not king does say is that people are willing to pay far more for point-to-point communication than for the famed content. That is likely to be reflected in what kinds of networks are built, and which companies succeed. It inverts the usual ordering of priorities, making point-to-point communication central, and content secondary. The fights over control of movie distribution may be a distraction from the main business of communication. As a simple example of what this may mean in practical terms, most broadband access links, such as cable modem and DSL ones, are designed to be asymmetrical, with higher capacity on the link to the home than to the network. The expectation is that these connections will be used primarily to pull content to the consumer. However, if the consumer places much higher value on personal communication than on content, the case for symmetrical connections becomes stronger. That may mean that fiber to the home may be justified sooner than expected. In the wireless arena, it suggests (as is explained in greater detail in Section 15) that for much of the next decade, the best strategy will be to emphasize regular voice transmission, supplemented by email and various low-bandwidth data transfers. Music and video services are likely to be delayed until much later.

Before continuing, it is worth considering a basic issue, namely, what is content? This word derives from the Latin “contentum,” which means “that which is contained,” but this derivation is not very descriptive. There is no precise definition, but generally content is used to denote material prepared by professionals to be used by large numbers of people, material such as books, newspapers, movies, or sports events. That is the sense in which it is used in this work. In general, content is distributed by “mass” or “broadcast” communications systems. Until a few decades ago, such services could be distinguished easily from “point-to-point” (or, more precisely, “person-to-person”) communications, which included first class letters and phone calls, and were specific to the people involved in the transaction. These two types of communications were sometimes combined during distribution, as in the postal system, which carried both letters and newspapers, in an early example of “convergence.” However, there was a noticeable distinction in how these two types of communication were prepared, handled, perceived by the recipients, and (a point discussed at great length already) in how much people were willing to pay for them.

During the last few decades, the distinction between point-to-point and broadcast communication began to blur. Computers allowed for the mass preparation of personalized letters offering credit cards, say. Answering machines and voice response systems led to machine-mediated point-to-point communications. Individuals were able to reach large audiences through postings to netnews, or, more recently, through their personal Web pages. We can expect this evolution of communications to continue, and eventually to achieve that convergence in which there will be a continuum between point-to-point and broadcast communication. However, we are not there yet, and won't be for a while.

In this work I do not classify information services such as weather, directory assistance and airline schedules as content. Many of the standard phone calls access just such services, and the Internet is leading to increasing usage of them. I also do not classify most of ecommerce as content. Somebody going to the Godiva Web site may be exposed to creative work in the ads flashed on the screen, but is interested in purchasing a tangible good. These types of interactions will flourish on the Internet, and some will be merging with content, but they are more typical of the standard point-to-point communications.

One of the main lessons from the history of communications is that technologies are often adopted rapidly, but seldom at the astronomical rates that popular imagination associates with the Internet. This applies even to the Internet, where change tends to be less rapid than is often thought. Browsers were adopted rapidly. The first informal release of Mosaic took place in the spring of 1993, and in under two years, the majority of Internet traffic was Web-related. However, that was an exception. Other changes have been slower. Just consider Internet telephony, introduced back in 1995. It is finally beginning to have a noticeable effect, but it is far from dominant. As another example, Amazon.com has had a striking impact on perceptions of ecommerce. Yet after more than four years, less than 10% of book sales take place online, and Amazon.com's investors are learning the virtues of patience. "Internet time" is a myth. In general, the time a new technology takes to become widespread has not changed much in the last half century.

A modern maxim says: "People tend to overestimate what can be done in one year and to underestimate what can be done in five or ten years."

(footnote on p. 17 of [Licklider])

Even technologies with compelling advantages tend to take a decade to dominate markets. Fax machines took about 10 years, from the introduction of the first inexpensive models until they became ubiquitous. Cell phones, one of the fastest growing industries, have taken about 15 years to reach their

present level (cf. tables 2.2 and 3.5). Cable TV has taken three decades to reach about 60% of U.S. households. Music CDs and more recently DVDs show similar patterns, taking on the order of a decade to reach dominance. For more examples and a discussion of rates of change, see [Odlyzko3]. Aside from the unusually rapid ascendancy of the browser-Web combination, fast change is usually associated with the presence of forcing agents. These can be either governments or a small number of key decision makers who can shift an industry's direction. An example of such forcing agents are the information technology managers at large banks and other enterprises. When they decided that mainframes were obsolete, they did so over a short period of time, and this led to a catastrophic decline in IBM's fortunes. A transition of voice phone traffic from circuit switched transmission to the Internet might occur rapidly for just this reason. The carriers might be able to implement it rapidly, since it could be done essentially invisibly to the end users, and the decisions would need to be made by only a few people.

The decade-long diffusion periods listed above for consumer goods and services are due to the inertia of the millions of people who have to individually decide to adopt a new technology. Most new products and services are in that category. Sociological changes are even slower, taking a generation or two. Normal change, with a simple shift in technology that offers serious advantages over older, more established competitor (as with CDs over vinyl records, which provided higher quality sound reproduction, or cell phones, which offered mobility, even at the cost of sound quality), takes on the order of a decade.

The speed with which new technologies diffuse has a direct relevance for the question of whether content might be king of the Internet. It is possible to make a case that even if content is not king now, it might be king in the future, when convergence does occur. Some evidence for this can be derived from the comparison of cable TV and cell phone industries. Their revenues and growth rates were cited as demonstrating that point-to-point communication is more important than content. However, there is another way to look at the data. As was mentioned, about a third of the cable TV revenues go to pay for content. What that means is that two thirds pay for the network. (According to some reports, carriers receive up to 90% of the revenues from some content, such as hard-core pornography pay-per-view movies.) If convergence moves delivery of most of the content to the Internet, and transport grabs two thirds of the total revenues from content, then the network will get a huge new source of revenue. Total content spending in the U.S. is comparable to that on the phone system, so two thirds of that would make a huge difference to network financing. This scenario is not totally implausible, since the current content distribution system is grossly inefficient. Book authors and musicians typically receive

in royalties less than a tenth of the price consumers pay for their creations. It is a striking observation that a participant in Amazon.com's affiliate program can sometimes get more money from a book sale generated by a link from his or her home page than the author does from royalties! The Internet offers a chance to reduce some of the inefficiencies of the current system. For the current dominant content producers, the real threat from the Internet is probably less from piracy, and more from disintermediation. New producers can come in and, unburdened by the high overheads and obsolete habits of established players, can offer better deals to both the creative talent and the consumers.

The scenario outlined above, in which Internet transport grabs the lion's share of revenues from content, is conceivable, but very unlikely. One argument against this scenario is based on a simple historical observation. No broadcast medium has ever been replaced by another; despite predictions to the contrary at various times in the past, newspapers were not killed by radio, nor radio by television. However, that argument may not be valid. The Internet is a disruptive technology, it does have unprecedented ability to emulate other delivery mechanisms, and we are already seeing rapid growth in music delivery on it. As an example of what can happen to even the most solid-seeming businesses, just consider the *Encyclopaedia Britannica*. (Its problems started even in the pre-Internet days, with CD-ROMs.) In spite of having the greatest brand name and by far the best content in the encyclopedia field, it has been floundering, and has yet to find a viable business model. Thus dramatic changes are indeed possible in the electronic environment. Still, there are other, more substantial, arguments against the content thesis.

The Internet will surely have a major impact on the content industries. However, as was discussed above, consumers are slow to change their behavior. Even the *Encyclopaedia Britannica* has had a decade in which to flounder. Couch potatoes are not going to abandon their TV sets right away, especially when computer and cable TV penetrations in the U.S. are under two thirds of households, and not growing rapidly. Even in the business environment, adoption of new technologies that require a thorough re-engineering of all internal processes is slow. The business-to-consumer dot-coms have discovered this already, and the business-to-business ones are in the process of learning. Although we are living in the Internet era, fundamental change is not all that much faster than a century and a half ago. The adoption of the telegraph by railroads, discussed in Section 13, did lead to huge efficiency gains, but it was slow as well.

Slow adoption of new technologies means that convergence will be spread over a decade or more, and there will be continuing competition from traditional media, as well as increasing diversity of delivery mechanisms for content. This may mean that the writers and artists will get a bigger share

of the pie. That appears to have been the trend over the last few decades, with movie actors and professional sport stars increasing their share of the revenues their work brings in. It is less certain whether carriers will manage to improve their share of the content pie to the same extent. There will certainly be a shift of revenue towards broadband services, but content distribution may not be the largest contributor to it.

The main reason to question whether content will ever make giant contributions to network costs is that by the time convergence is likely to occur, at least a decade into the future, content transmission is likely to be a small fraction of total traffic. Further, most content will probably be distributed as ordinary file transfers, not in real-time mode. The various rates of growth that contribute to these predictions are discussed later. Right now we note that if these predictions come true, then it will be hard for networks to charge much for the transport of content. High prices could be charged for content distribution if content made up most of the traffic, or else if content required special transmission quality (such as that needed for real-time streaming traffic). Since neither of these conditions is likely to be satisfied, though, content will probably constitute just some of the huge number of large files, many encrypted, that will be flying around the network. How could carriers pick out the content files for special pricing?

Let us next consider the predictions for Internet traffic mentioned in the preceding paragraph. I will treat them briefly, and refer the reader to [CoffmanO2] for more detail. (There is some discussion of this point in Section 4 as well.) Internet backbone traffic appears to be about doubling each year. Further, advances in photonics and additional fiber deployment appear to allow for a doubling of network capacity each year for the next decade. In early 2000, Internet backbone traffic in the U.S. appeared to be less than a third of voice traffic in volume, but is likely to become larger by about 2002, with the transition point slightly later for the rest of the world. However, if voice traffic were to be packetized, it would almost surely be compressed, and then its volume would already be less than that of Internet traffic (at least in the U.S.). Broadcast TV would still overwhelm the Internet of the year 2000, but at 100% annual growth rates, it will not be too long before there is more than enough capacity to provide a high quality video channel for every person on Earth. This is because bandwidth that is likely to be required to satisfy any single person's real-time transmission needs will not be increasing fast.

The versions of Moore's Law that hold in different industries operate at diverse rates. Microprocessors are doubling in computing power every 18 to 24 months, while fiber transmission capacity is doubling every year. On the other hand, display technology is advancing extremely slowly. Broadcast TV resolution has been static for several decades, and even the planned move to HDTV will require only a modest increase in bandwidth. Thus satisfying the needs for real-time multimedia transmission

will not require much of the Internet's capacity.

Content is likely to form only a small fraction of Internet traffic for reasons explained above. In addition, real-time delivery of content is likely to be an even smaller factor, for reasons discussed in much greater detail in [CoffmanO2]. Transmission capacity is approximately doubling each year, which is a much faster rate of improvement than Moore's Law for semiconductors. However, hard disk storage capacity is also about doubling each year. Furthermore, that capacity is already huge. At the beginning of 2000, the U.S. Internet backbones appeared to be carrying about 12,000 TB (terabytes) of traffic each month. However, the total world hard disk capacity was 3,000,000 TB. Thus it would take about 20 years to transmit all the data on those disks over U.S. Internet backbones. Nobody proposes to do that (and why would anyone want to send all the duplicate copies of Windows 98 around, in any case?), but this comparison helps in visualizing the technology landscape. The presence of huge local storage capacity in local PCs or cable TV setup boxes will make it much more attractive to send even content as files, not as real time transmissions. There will be a growing volume for real time multimedia traffic, for applications such as videoconferencing. However, such applications are likely to be swamped by ordinary file transfers. The dominant mode of operation is likely to be fast (eventually much faster than real time, but initially often slow) download to local storage, fast transfer to whatever display device one wishes to use (often a mobile information appliance), and then playback. That is already the model we see emerging with MP3, Napster, and TiVo. The advantages of this model include the ability to implement it now, before the Internet can be made ready for real-time streaming media. It also accommodates gracefully the forecasted explosive growth in small mobile devices, which will often have small storage and low bandwidth over wireless links, and thus will be most useful if they can get data from local storage. This model also allows for easy integration with special hardware for intellectual property protection.

That real-time multimedia traffic would not dominate the Internet has been predicted several times in the past. It is an obvious conclusion from the rapid increase in traffic. Already the authors of [deSolaPITH] noted that in the early 1980s data traffic was growing much faster than voice communication. They observed that if that trend continued, eventually most transmissions would not be seen by human eyes nor heard by human ears. In a similar vein, in discussing general digital data volumes in 1997, Michael Lesk predicted that "the typical piece of information will *never* be looked at by a human being" [Lesk]. Bill St. Arnaud appears to have been the first one to predict in the specific context of the Internet that the general expectations for streaming multimedia domination were unlikely to come true [StArnaud, StArnaudCFM]. Further arguments were presented in [Odlyzko8].

The huge volume of local storage will surely stimulate the generation of non-content traffic. Both corporations and individuals so far have had no difficulty filling their disks with data. We can expect this to continue, although predicting the exact source of that data is uncertain. For firms, various databases will likely continue to proliferate and grow. For residential users, pictures are the leading candidates for filling those disks. Ease of use, lower cost, and instant gratification all stimulate use, and digital camera owners appear to be taking many more pictures than they ever did with regular film. The same electronic technology that is producing better disks and processors is also producing better cameras. Historically, it appears that privately taken pictures have traditionally been the dominant source of data. An interesting accounting of all the information stored in the world in 1997 by Michael Lesk [Lesk] found that home photographs were the dominant component. They contributed about 500,000 TB each year (even when one assumes that each picture is stored as a modest 10 KB JPEG file). By comparison, all the texts in the Library of Congress amounted to around 20 TB, while the graphics and music in that collection came to about 3,000 TB. Thus even this great library contained less than 1% of the world's information. (The publicly accessible Web pages currently contain a few tens of terabytes, just a few percent of what the Library of Congress has, but comparable to the text collections in that library.)

An obvious comment to the estimates above is that the purpose of a library is to select the most valuable material, and that most of those photographs contributing to the 500,000 TB are of no interest to most people. That is true, but that does not stop those pictures from being taken, and it will not stop an explosion in volumes of data collected this way in the future. A few pictures or video clips will turn out to be of great interest, in spite of amateur production. Just think of the Zapruder film of the JFK assassination, or the Rodney King video. More importantly, many of the pictures being taken are of interest, or might be of potential interest, to at least one person. Most of the world will have no interest in a picture of your newborn baby, but your mother will cherish it. Similarly, in the future you will be taking digital video clips of your children and sending at least some of them to your mother. Many of the video shots will intentionally be made with the hope of not having to see them, as with security monitors. (Note that some of the earliest applications of miniature taperecorders and video cameras has been to snoop on child care providers. There are obvious privacy implications of current and future camera technologies that are not pursued here, but are discussed in [Garfinkel], for example.)

Other examples of data that may be filling our disks are suggested by the entry in Table 2.1 for sporting goods. Spending in this area comes to about half of consumer spending on content, showing how highly these products are valued. One can easily imagine future generations of body and equipment sensors that would record precisely all details of a player's movements in tennis, say. These details

would then be fed into systems that would analyze the motions, compare them with previous games, and produce high quality graphical displays to help improve the player's game. There is practically no bound on the amount of data that could be generated this way.

The data that will be generated is likely to be shared using programs descended from Napster. Email and the Web may not be flexible enough. Napster is currently attracting huge attention because of the threat it poses to conventional music distribution channels. However, that may turn to be less important than its ability to facilitate sharing of files. Napster itself is too limited, as it is designed to handle just MP3 music files, and is also centralized. Yet it has already inspired creation of tools such as Gnutella, which are much more general and decentralized. Given the growth of local storage, and the increasing availability of tools to fill that storage with video clips and other material, it is possible that tools like Gnutella may become more important to the Internet than the Web.

The discussion above is futurology. We cannot be certain how the Internet will evolve. However, history teaches us several lessons, which are discussed in greater detail in later sections. One is that the growing storage and communication capacities will be used, often in unexpected ways. (For a careful study of the many early predictions about the future of the telephone, and the actual evolution of that service, see [deSolaP1, deSolaP2], for example.) Another important lesson is that the value of the myriad social interactions has often been underestimated. Only a tiny fraction of the information passing through communications systems has ever been high quality scholarly knowledge. Even in more prosaic transmissions, we have moved from Samuel Morse's solemn "What hath God wrought?" to Alexander Graham Bell's utilitarian "Mr. Watson, come here, I want you," to the banal "How was your lunch?" that is so common today. The volume of communication has increased, the importance of a typical message has decreased, and the attention we pay to such a typical message has decreased. However, the aggregate value of all these exchanges has increased.

As is described in more detail in later sections, sociability was frequently dismissed as idle gossip, and especially in the early days of the telephone, was actively discouraged. For example, a 1909 study of telephone service commissioned by the city of Chicago advocated measured rate service as a way to reduce "useless calls" [JacksonCW]. Yet the most successful communication technologies, the mail and the telephone, reached their full potential only when they embraced sociability and those "useless calls" as their goal. That seemingly idle chit-chat not only provided direct revenues, but it encouraged the diffusion of the corresponding technology, and made it more useful for commercial and other applications. Such social interaction frequently function to grease the wheels of commerce.

This work is largely the story of the development of the communications infrastructure from the point of view of the user. Usually an infrastructure is noted for being unnoticed; it is simply there, something we come to rely upon, do not have to think much about, and are horribly inconvenienced when it malfunctions. Electricity, water, mail, and the phone are excellent examples. Yet they all took much effort to reach this stage. The key failing of the telegraph, as is detailed in later sections, is that it never became a true infrastructure component. It was a revolutionary technology, the “Victorian Internet,” as one writer has called it [Standage], but it was too cumbersome and too expensive to attract much usage, and in particular never carried much of that “useless” social traffic that pushed the mail and the telephone to their eminent positions.

Although social uses are important to the telephone industry, a glance at Table 2.1 shows that most of the money comes from businesses. Consumer spending on phone service brings in only about a third of the total revenues. (The figures for total revenues, \$256 billion in 1997, and consumer spending, \$85 billion, come from different sources. It is possible that consumers spend somewhat more, especially for cell phones, than is reported in the \$85 billion figure. However, even if one makes the most likely adjustments, it still appears that business spending on telephony is far larger than that of households.) That has been the historical trend, and many communication services, including the phone, were initially devoted almost totally to business uses. Traditionally, commercial users have subsidized residential ones. Sometimes this was done even involuntarily, as in higher rates dictated by carriers or by government regulators, and sometimes voluntarily, as in paying for toll-free 800 numbers. One of the purposes of this work is to examine the history of such subsidies. It appears probable that they will also play a large role on the Internet. We may very well end up with a system in which the largest monetary contribution will come from commercial users, the second largest for households paying for point-to-point communication, and the smallest by the transport component of charges for content.

The value of a broadcast network is usually regarded as proportional to the number of users in it. On the other hand, a point-to-point communication network is often said, by Metcalfe’s Law, to have value proportional to the square of the number of member. This then leads to the conclusion that eventually, once a single network like the Internet reaches a large enough size, point-to-point communications will provide much higher value than broadcast. There are some problems with this argument. In particular, Section 4 shows, using historical evidence as well as general principles, that Metcalfe’s Law does not reflect properly several other important factors that go into determining the value of a network. However, the general thrust of the argument and the conclusion are valid. Certainly all the historical

evidence cited throughout this work supports the conclusion that point-to-point communication is what matters the most.

Whether content is king or not has direct relevance for the question of whether the Internet will continue to be an open network, or whether it will be balkanized. If content were to dominate, then the Internet would be primarily a broadcast network. With value proportional to the number of users, there would be few inherent advantages to an open network. The sum of the values of several completely or partially separate networks would be the same as of a unified network. On the other hand, if point-to-point communications were to dominate, and if Metcalfe's Law were to hold, there would be strong economic incentives to a unified network without barriers. This is considered more fully in Section 4. The general conclusion there is that even though Metcalfe's Law is not fully valid, the incentives to maintain an open network are likely to be very strong. This will be largely because content is not king, and effective point-to-point communication will demand easy interconnection.

An extreme form of the "content is king" position, but one that is shared by many people, and not just in the content industry, was expressed recently by the head of a major music producer and distributor:

What would the Internet be without "content?" It would be a valueless collection of silent machines with gray screens. It would be the electronic equivalent of a marine desert - lovely elements, nice colors, no life. It would be nothing.

[Bronfman]

The author of this claim is facing the possible collapse of his business model. Therefore it is natural for him to believe this claim, and to demand (in the rest of the speech [Bronfman]) that the Internet be architected to allow content producers to continue their current mode of operation. However, while one can admire the poetic language of this claim, all the evidence of this paper shows the claim itself is wrong. Content has never been king, it is not king now, and is unlikely to ever be king. The Internet has done quite well without content, and can continue to flourish without it. Content will have a place on the Internet, possibly a substantial place. However, its place will likely be subordinate to that of business and personal communication.

3. Growth in communications: quantity and quality

The explosive rise of the Internet is only the most recent chapter in a remarkable history of humanity becoming increasingly connected, the "annihilation of space and time," in a phrase going back at least to

the early 19th century. The volume of messages addressed individually to each one of us has increased about a thousand-fold over the last two centuries. Much of this growth has been slow, and accomplished through the power of compounding. However, growth rates rivaling that of the Internet have been observed occasionally before, for example in the development of the telegraph. The wireless industry is experiencing similar growth even now.

This section surveys the long trend of increasing communication. The first part is devoted to quantitative measures, the growth in spending and in volume of transactions. The second part considers the qualitative changes in how we approach communications. The role and evolution of pricing are then the subject of the rest of the paper.

The communications technologies considered in this paper are primarily point-to-point ones, and generally exclude mass media, such as newspapers, book publishing, radio, and TV. One reason is to keep the size of this work manageable. A related one is to avoid the complexities of trying to measure the volume of information delivered by broadcast media. With point-to-point communication, it has often been true that the entire content was read, heard, or seen by the recipient. With broadcast communication, accounting is harder. If 70 channels come across on a cable TV coax, should one measure the information on all of them all the time, just when the TV set is turned on, or just the channel that is watched when the set is on? Should one also consider the fraction of the attention that the people in front of the set are devoting to the screen, as opposed to talking to each other? It is not easy to come up with a fair measure. (See [deSolaPITH] for an attempt to do this. Also note that, as was indicated in Section 2, similar problems will increasingly arise on the Internet, as a diminishing fraction of the traffic passing through it is seen or heard by people.) The third, and most important reason for concentrating on point-to-point communications is that, as was shown in Section 2, it is far more important than broadcast.

The communications industry as a whole has been expanding for centuries. Not only that, but most of the prominent services have continued growing. The tables in this section provide more eloquent testimony to this historical fact than words can. Even the venerable mail, often derisively called “snail mail,” is still growing. The remarkable story of this ancient yet still widely used technology is summarized briefly in Table 3.1, which shows key statistics in the development of the U.S. Postal Service (USPS). (For most of its history it was known simply as the U.S. Post Office.) Although there have been many predictions that mail would decline as a result of competition from the telegraph, the telephone, and more recently from email, it has continued to expand. It went from fewer than a million pieces of mail in 1790 to almost 200 billion in 1998. The mix of traffic has changed. It was about half first class

Table 3.1. Growth of the U.S. Postal Service

year	expenditures (millions)	expenditures as percentage of GDP	pieces of mail (millions)	pieces of mail per person per year
1790	\$0.032	0.02%	0.8	0.20
1800	0.214	0.05	3.9	0.73
1810	0.496	0.09	7.7	1..07
1820	1.161	0.18	14.9	1.55
1830	1.933	0.21	29.8	2.32
1840	4.718	0.28	79.9	4.68
1850	5.213	0.20	155.1	6.66
1860	14.87	0.39		
1870	24.00	0.33		
1880	36.54	0.35		
1890	66.26	0.51	4,005	63.7
1900	107.7	0.58	7,130	93.8
1910	230.0	0.65	14,850	161
1920	454.3	0.50		
1930	803.7	0.89	27,887	227
1940	807.6	0.81	27,749	211
1950	2,223	0.78	45,064	299
1960	3,874	0.77	63,675	355
1970	7,876	0.81	84,882	418
1980	19,412	0.70	106,311	469
1990	40,490	0.70	166,301	669
1998	57,778	0.68	197,943	733

mail, and half newspapers in the early 19th century, if we count separate pieces delivered. It is about half first class and half “junk mail” today, and even a large fraction of the first class mail is targeted advertising. (It is estimated that only about a tenth of the first class letters are sent by households, see for example [Wolak].) Various new services have been added, such as money orders, home delivery, and parcel post. Even without the new services, the volume of letters alone has been increasing. Overall, total mail volume and spending have grown at impressive rates. Prices, discussed in Section 12, have decreased. The decline in prices was dramatic in the 19th century. Even in recent years prices have been decreasing, if we adjust them for increases in earnings. Moreover, this was not an isolated case of a frontier economy getting civilized. The French postal system in the 19th and early 20th centuries had similar growth patterns, as can be seen in Table 3.2. The increasing role of communication was common to all countries as they industrialized. (Table 12.1 shows data for the British mail system that behaved similarly, although there were major differences that are discussed in detail in Section 12.

Analogous trends can also be seen in the statistics of the Swiss postal system presented in Table 21.1.)

Table 3.2. Growth of the French postal system

year	pieces of mail (millions)	pieces of mail per person per year
1830	104	3.2
1840	147	4.3
1850	254	7.1
1860	444	11.9
1870	733	18.8
1880	1,231	33.0
1890	1,763	46.3
1900	2,433	63.4
1910	3,758	96.1
1920	4,162	108.1
1930	6,281	153.2
1940	4,354	105.7
1950	4,050	98.1
1960	6,093	133.9

USPS revenues went from 30 thousand dollars (in current dollars, not adjusted for inflation) in 1790 to almost 60 billion today, and are still increasing. Such nominal dollar measures ignore the effects of inflation and economic growth. Hence it is more informative to compare the postal system to the whole economy. We find that spending on mail in the U.S. grew from 0.02% of the Gross Domestic Product (GDP) in 1790 to almost 0.9% in 1930. Since then, it has declined to about 0.7%. However, this decline is misleading, since the statistics in Table 3.1 do not include competing private delivery services. UPS has taken the lion's share of the parcel business from the USPS. (Parcel delivery does not count as communication by the standards of this paper, but it is hard to disentangle it from regular mail. Parcel post was introduced by the USPS in 1913, and contributes significantly to its budget, although not to the extent it did before UPS made its inroads.) The annual revenues of UPS are around \$27 billion. Further, express delivery companies such as FedEx, DHL, and Airborne Express are largely in the message communications business. FedEx, the largest by far, has annual revenues around \$17 billion, with about \$12 billion of that from U.S. traffic. If we added up the contributions of all these express services, we would likely find out that the fraction of U.S. GDP devoted to basic mail services is still growing. The USPS, like other countries' postal systems, and even the express delivery companies, does face challenges, particularly from the growth of email and the Internet in general. Those challenges also represent opportunities, though, as ecommerce still requires physical delivery of

most goods.

Table 3.3. U.S. domestic telegraph industry

year	revenues (millions)	revenues as percentage of GDP	messages handled (millions)	messages per person per year
1870	\$6.7	0.11%	9.2	0.23
1880	10.6	0.09	29.2	0.58
1890	20.1	0.15	55.9	0.89
1900	22.8	0.12	63.2	0.83
1910	30.7	0.09	75.1	0.81
1920	124.4	0.14	155.9	1.46
1930	148.2	0.16	212.0	1.72
1940	114.6	0.11	191.6	1.45
1950	178.0	0.06	178.9	1.18
1960	262.4	0.05	124.3	0.69
1970	402.5	0.04	69.7	0.34

Postal system revenues are more than three times larger as a fraction of GDP than they were 150 years. In spite of this, mail is no longer the dominant communication technology, as it is dwarfed by the phone industry. Comparison of tables 3.1 and 3.4 shows the telephone passed the mail in revenues by 1920, and today is about four times as large. Adding the contributions of these two industries, we find that their role in the economy, measured as a fraction of GDP, has increased almost 20-fold over the last 150 years, from 0.2% to over 3.5%.

Table 3.4. Development of U.S. phone industry

year	revenues (millions)	revenues as percentage of GDP	phone calls per day (millions)	phones calls per person per day
1890	\$16	0.12%	0.0015	0.00002
1900	46	0.26	7.8	0.1
1910	164	0.46	35.6	0.4
1920	529	0.58	51.8	0.5
1930	1,186	1.32	83.5	0.7
1940	1,286	1.29	98.8	0.7
1950	3,611	1.27	171	1.1
1960	8,718	1.73	288	1.6
1970	19,445	1.99	494	2.4
1980	59,336	2.13	853	3.8
1990	133,837	2.33	1,272	5.1
1998	246,392	2.88	1,698	6.3

The tide of increasing communications volumes has been powerful enough to lift most services. The telegraph did not stop the mail from growing. The telephone took a long time to eliminate the telegraph. Wireless telephony and the Internet have been growing explosively, yet wired telephony is still growing. Even fax use is increasing. Yet if anything should have been eliminated by the Internet, it surely should have been the fax. Most messages sent by fax are typeset on a computer, printed out, and then fed into a fax machine. Why not eliminate some of the redundancy, and send the message via email? Well, that is beginning to happen, as is faxing directly from a computer. Still, the fax is ubiquitous, inexpensive, easy to use, and provides some features, such as signatures, that the Internet can and will eventually provide, but does not offer now.

Table 3.5. Growth of U.S. cell phone industry

year	number of subscribers (millions)	revenues (millions)
1985	0.20	\$ 354
1986	0.50	667
1987	0.89	942
1988	1.61	1,558
1989	2.69	2,480
1990	4.37	4,061
1991	6.38	5,076
1992	8.89	6,688
1993	13.07	9,009
1994	19.28	12,592
1995	28.15	16,461
1996	38.20	21,526
1997	48.71	25,575
1998	60.83	29,638
1999	76.28	37,215

A technology that survived for longer than the fax (at least so far) is the electric telegraph. One can speculate that had the telephone not been invented, the telegraph might have spread more widely and become less expensive. Eventually it might have developed into something like our Internet, going from digital transmission directly to digital, bypassing the analog phone stage. There was intense development of new applications of telegraph technology in the second half of the 19th century for stock tickers as well as fire and home alarms [TarrFG]. However, these developments came to nought. The telephone started encroaching on the bread and butter business of the telegraph, and the latter could not compete effectively. Still, as Table 3.3 shows, usage of the telegraph kept growing almost to

the middle of the 20th century. (The actual peak year for sending domestic telegrams in the U.S. was 1945.) Had Table 3.3 combined statistics for telegraph and telex traffic (since telex was to a large extent a substitute for the telegraph, and drove the latter into the ground), the catastrophic decline in volume of messages that is visible in the data would have been postponed. Instead of showing up during the 1950s and 1960s, it would have been visible only in the 1980s. It was only the fax that provided a way for the phone system to emulate telegraph and telex services, and eliminate those.

The long survival of the telegraph (and the fax) suggests many lessons about rates of technological change. One that is particularly relevant for this paper is that efficiency often plays a minor role when relatively inexpensive goods or services are involved. The Internet can emulate other communication technologies much better than the phone could emulate the telegraph, say. Still, the power of inertia should not be underestimated. After all, why hasn't the Internet eliminated the fax yet?

Where did the steadily increasing spending on communication that is documented in the tables come from? The general progression for new technologies has been to start as an expensive and limited service. Then, as a result of technological improvements as well as economies of scale, prices usually declined and utilization increased. Growth in usage was accomplished by development of new types of usage. For some very expensive services, governments were the first customers. As Samuel Johnson said, "when a man knows he is to be hanged in a fortnight, it concentrates his mind wonderfully." The French revolutionary government of the early 1790s was fighting for its life, and did face the prospect of being hanged (or perhaps guillotined, were its enemies to turn its own invention on it) should it lose the war. This undoubtedly increased its willingness to fund Chappe's optical telegraph, mentioned briefly in Section 13. Even earlier, mail systems started out in most cases as government communications services.

Some communication technologies, such as the electric telegraph and the telephone, went into public service right from the start. Initial usage was typically limited to special high-value commercial communications, since prices were high. In the Introduction, the sending of a letter in early 19th century England was compared (on the basis of earnings) to the purchase of a book today, costing about two hours of work. Table 13.1 shows that the first reliable commercial telegraph communication between New York City and London, established in 1866, cost \$100 for 10 words. That was several months' wages for a laborer then, comparable to the work time needed currently to pay for a round trip on the Concorde supersonic plane. By 1880, prices had decreased 20-fold, so that relative to earnings, the cost of a similar 10-word telegram became comparable to an exceptionally good airplane excursion fare from New York to London today. Thus prices were still very high, and careful consideration of

the need for each telegram was called for. Services that expensive could only be afforded for social purposes by the very rich. Otherwise they were restricted to commercial purposes, where the savings justified the expenditures. Over shorter distances, prices were lower, but still high compared to either mail or earnings, and so usage was also low, and restricted to urgent needs. As Table 3.3 shows, at the end of the 19th century there wasn't even one domestic telegram per person per year in the U.S.. (Today there is slightly more than one airplane round trip per person per year.)

Telephone service initially was also extremely expensive. For example, a century ago, unlimited local service in New York City cost \$20 per month. That is comparable to \$2,000 per month currently, when one considers prices in relation to average earnings. How many people would be willing to pay \$2,000 per month for Internet connectivity today? That such prices were paid testifies to the value of communication services as perceived by users. The precise economic benefits of communication are a complicated subject beyond the scope of this work. However, they have consistently been high, often demonstrably so. (At a high level, this claim is proved by the statistics in the tables throughout this work. Since nobody was overtly forced to pay for any of these services, yet usage and spending went up, customers must have been getting their money's worth.) As one example, the first transatlantic cable to carry traffic was completed in 1858. It worked only sporadically, and broke down completely after a few weeks. It never carried commercial traffic. However, when the news that the Indian Mutiny was being brought under control arrived in London, the British government was able to use it to send a telegram to stop the departure of two regiments of Canadian troops for India. That single message was claimed to have saved \$250,000. Similar, although usually smaller, savings were often attained in commercial settings. As a simple example (others will occur throughout this work), the first telegraph connection to Australia was opened in 1872. "In Sydney the softgood firm of Prince, Ogg and Co., doing business of between [\$2,500,000] and [\$3,000,000] a year, reduced the value of stock held on the premises from [\$1,300,000] to [\$500,000] once they could telegraph to London for supplies" [Inglis].

The main reason telecommunications services are being liberalized and privatized around the world is to spur innovation and greater investments. Many careful studies have been conducted that documented the contribution of communications to speeding up economic growth. These benefits are easily appreciated by individuals. We read stories of farmers in Bangladesh renting time on cell phones to get the most recent market prices. A century ago, the highest penetration of phone service in the U.S. was in the Midwestern farming states. Even when there were no direct monetary savings, the savings in time from not making unneeded trips or arriving right on time were immense, and easy to understand for anyone.

Once mail and phone service prices became low enough, social uses exploded (a topic mentioned already in Section 2, and treated in greater detail in Section 4). This enabled those two industries to increase their shares of the economy. The telegraph never moved beyond the high-priced and infrequent commercial use stage, and eventually died.

The large and often immediately noticeable economic effects of communications meant that any changes in availability or pricing of services could have a large effect on the economy. Even regular mail played such a role. Much of the maneuvering in the U.S. Congress in the first half of the 19th century over postal rates had to do with the competitive positions of newspaper publishers in small towns versus those in large ones, and those in the South versus those in the Northeast. (See [John1, Kielbowicz2] for details.)

The interaction of mail with economic life developed gradually. This was largely because mail evolved gradually, and was a familiar service, simply performing better and less expensively what everyone could do, namely carry letters at the speed people could travel. Postal services did play a major role in the economy even in the early 19th century (as is visible in tables 3.1 and 12.1, and is discussed in greater detail in Section 12). However, mail was apparently not associated in the public eye with dramatic changes, at least not negative ones.

The electric telegraph introduced a novel phenomenon. The Internet is sometimes portrayed as a uniquely potent disruptive technology. Yet it is worth keeping in mind that the electric telegraph had a similar effect. In many ways, the telegraph was a greater innovation than the Internet. It was the first technology that effectively separated communication from transportation. It had a dramatic effect on science and technology, and it facilitated huge economic changes.

The role of the telegraph in transforming the economy of the 19th century can be illustrated briefly by considering the history of the Rothschild banking dynasty [Ferguson]. The Rothschilds were definitely not technophobes. They were the world's largest financiers of railroads in the middle of the 19th century, and much of their wealth derived from that source. On the other hand, they were hostile to the telegraph, since it eroded their competitive advantage. In the first half of the 19th century, the Rothschilds attained unprecedented dominance of international finance because they alone had a cohesive organization covering all the major financial centers of Europe. The closely knit family could trust each other, and in addition they had a private communication system that provided them with news in advance of rivals. These advantages were much more important to Rothschilds' pulling ahead of their competitors, such as the Barings, than any early coups during the Napoleonic wars. The telegraph threatened their dominance.

Throughout the 1850s, [James de Rothschild] repeatedly complained that “the telegraph is ruining our business.” The fact was that the telegraph made it much easier to do what the Rothschilds had managed so ingeniously before, namely to conduct financial business between affiliated houses over long distances. ... “It appears,” James complained in April 1851, “that yesterday a great many German scoundrels sold [French] railway shares in London with the telegraph...Since the telegraph became available, people work much more. Every day at 12 they send a despatch, even for trivial deals, and realise [their profit] before the bourse closes the same day.” Once, the Rothschilds had been able to steal a march on their rivals with their unrivalled system of couriers and carrier pigeons; but now “anyone can get the news.” James could see that there was no alternative but “to do the same,” but it still struck him as a “crying shame that the telegraph has been established.” ... Such complaints were still being echoed by James’s son as late as the 1870s: although the Rothschilds had no option but to make use of the new technology, they always regretted the way it tended to broadcast financial news...

pp. 64-65 in [Ferguson]

The effect of the telegraph on the Rothschild banking business was an example of Schumpeterian “creative destruction.” New competitors sprung up, eroding the Rothschild’s competitive advantage and profit margins, and leading to greater economic efficiency.

Most businesses in the mid-19th century had a choice of using the telegraph or dying, and using it effectively often meant large changes in modes of operation. This new technology was closely associated with new industries, especially the railroads, and new styles of management. It did lead to much more centralized control in both government and business affairs.

The establishment of a direct telegraph line between England and India in 1870 was an event of far-reaching importance. The delay in communication was a great advantage to the Government of India in so far as it of necessity left the initiation of policy in urgent matters to its own hands, and enabled it to confront the Secretary of State with accomplished facts. But all this was bound to change when the Secretary of State had to be kept constantly informed of the course of events in India, and was in a position to issue immediate orders. Henceforth the Secretary of State exercised a far more effective control over the administration of India than was the case before, and the Viceroy really tended to be a mere “agent” of the Secretary of State.

For the role of the telegraph and the telephone in shaping commercial management operations, see [Chandler].

There is not doubt that the telegraph and the telephone had a great role in transforming society. However, many of the claims of dramatic changes caused by new communication technologies should be treated with caution. There were changes, but how big were they, and to what extent were they offset by other changes? Corporate management was centralized, but there was an explosion in ranks of corporations, so some general measure of autonomy by managers may well have increased.

Many communication technologies led to predictions that they would lead to a new era of universal brotherhood and peace. As we know, that did not happen with the telegraph and the telephone, and is unlikely to happen with the Internet. New institutions did arise. For example, the International Telegraph Union (which has since transformed itself into the International Telecommunications Union, the ITU), was the first permanent international technical organization. (The Postal Union followed later, since the urgency of standardizing mail communication was not as great as that of the much faster telegraph signals.)

There have been many predictions that the Internet would doom large corporations. After all, the generally accepted theory, due to Coase [Coase1], says that firms exist to minimize transaction costs. In the absence of such costs, the factory worker, the garbage collector, the secretary, the software developer, and the salesman could just contract for their contributions to any given job in a decentralized market, without the need for a cumbersome bureaucracy. Since the Internet reduces transaction costs, it has often been predicted that it would eliminate the need for large organizations, and reduce the world to cottage enterprises. However, while some large companies are divesting units, and there is a continuing growth of small new companies, we are also experiencing the greatest wave of mergers and acquisitions in history. Coase himself has explained the reason for this mixed picture.

If transaction costs were the only factor, then we would [become an economy of individual entrepreneurs]. But the fact of the matter is the cost of organizing is decreasing too. The question is whether the costs of transacting decreases as fast as the costs of organizing. My guess is that sometimes it does and sometimes it doesn't.

[Coase4]

The same phenomenon occurred in the second half of the 19th century, facilitated by the lower transaction and coordination costs offered by the telegraph. There does seem to be a strong tendency today

to move away from vertical integration. That mode of operation appears to be replaced by enterprises concentrated in narrower horizontal layers of the economy, but operating on a worldwide scale. Similar tendencies can also be discerned in the 19th century and early 20th centuries, facilitated by the telegraph and the telephone. This might seem absurd, especially in view of Henry Ford's vertically integrated River Rouge plant. However, even that enterprise was not anywhere near as vertically integrated as say the British and Dutch East India companies. Those ran shipping lines, plantations, banking, and sometimes entire countries or provinces. The general trend since then has been towards greater specialization.

There are persistent fears that the Internet will homogenize the world's societies, turning them all into slight shades of the American culture. Such fears are not new, since they were also associated with the telephone, and later with radio and television. Yet this has not happened, and linguists are even discovering that regional accents in the U.S. are diverging.

The Internet has stimulated a series of sociological studies, some of which (e.g., [KrautLPKM]) claim that it decreases human contact, while others (e.g., [Kanfer]) come to the opposite conclusion. This is just what happened with the telephone (see [deSolaP1, Fischer2]). It is certainly true that proportions of different types of interactions have changed. It appears hard to categorize them easily, though.

For better or worse, I expect these changes to facilitate a continuing transformation away from interaction in solidary communities and workgroups and towards interaction in far-flung, sparsely-knit and specialized social networks.

[Wellman]

There has been a measurable growth in physical distances over which we interact, facilitated by improvements in both transportation and communication. However, we do not have a "death of distance" in its extreme form, where we could live anywhere we wished and do any kind of job, and are unlikely to reach such a state. Sections 21 and 22 have further discussion on these topics.

The general conclusion is that it is hard to make any categorical statements about communications having caused any great social or economic transformation. New technologies have often facilitated changes, but they were seldom the sole factors. Precise predictions have seldom been correct. Society has adopted the tools that were offered to it in a variety of often unexpected ways. The one factor that is undeniable, though, is that apparent from the tables in this section. Historically communication has been extremely highly valued. Individuals and enterprises have been devoting increasingly large

fractions of their resources to it, and consuming rapidly rising volumes of it.

The growth in volume of information has led to a change in how we react to it. Over the last two centuries we have moved from scarcity to surfeit. There have always been some complaints about too much information. The Bible says that “[o]f making many books there is no end, and much study wearies the body” (Ecclesiastes 12:12). John Winthrop’s heavy correspondence from Massachusetts with many people in England led Richard Saltonstall to remark in 1632, “I fear he hath too many letter” (p. 217 of [Cressy]). One of the complaints that James de Rothschild had about the telegraph in the 1850s was that it generated too much work. “[The telegraph] meant that even when he went to take the waters for his summer holiday, there was no respite from the business: ‘One has too much to think about when bathing, which is not good’” (p. 65 in [Ferguson]). However, on the whole such reactions tended to be rare. People used to be information-deprived, and eager for any scraps they could get. The books [John1, Kielbowicz2] have many illustrative examples of the eagerness with which mail was received in 19th century America. A great visual testimony to the hunger for current information is provided by a photograph reproduced on p. 64 in [John3]. It shows the Post Office in Rochester, New York, on a Sunday in 1888, full of businessmen picking up their mail and sharing information.

Today, our task is to cope with the flood of information. There are many worries that restrictive copyright laws will limit our access to data. However, in practice what we see is an unprecedented abundance of information. With reduced marginal costs of distribution, content providers are fighting for eyeballs. Until a few years ago, you had to pay for a subscription to the New York Times, and (if you lived away from the New York City area) also had to wait, often a long time, for your copy to be delivered. Today, you can get it for free on the Web soon after the editors decide on the final layout. The New York Times is even soliciting people to allow it to send them email notifications of the latest editions! The main need is for tools for selecting from the flood of available information.

When communication opportunities were rare, and transmission slow, it was natural that great care was taken with correspondence. For example, in the British colonies in North America, it was usual for writers to compose a letter to fill most of the single sheet, but leave space for additions. This was natural given the delays in transmission, illustrated by the following:

John Hull commenced one letter on 12 December 1671, finished it on 31 January, but then had another couple of weeks in which to augment it. The reason was that ‘Mr. Greenought was ready to sail [on 9 February] but since it [the weather] came in very sharp and hath stopped him hitherto.’ The ship, with its basket of letters, did not leave Boston until 19 February.

Those who lament the lack of style in current letters compared to the often essay-quality compositions of the 17th or 18th centuries need to realize the different environment we operate in. We do not have weeks to compose a letter, and speed is of the essence. This trend is exacerbated with email. Email messages are often sadly deficient in style, spelling, punctuation, and grammar. Instant messaging is typically even worse. However, when it is necessary to deal with scores of email messages per day, it is natural to treat them as informal conversations. After all, are we expected to always speak in grammatically correct sentences?

New communication technologies require new modes of acceptable usage. Today people complain about rudeness of cell phone users in restaurants or on airplanes, and wonder at the strange sights of cell phone users with headsets who seem to be talking to themselves. The phone also required development of new rules of etiquette [Fischer2]. Even the wider use of mail for social communication led to “a burgeoning market for how-to manuals to teach ordinary Americans the once-arcane custom of maintaining a correspondence with distant friends and family” (p. 161 of [John1]).

Do we need this flood of messages? Complaints have been frequent. We already noted James de Rothschild’s gripe about the telegraph, and there were many more of this nature when the telephone arrived ([deSolaP1, Fischer2]). Today, we hear complaints about the overwhelming flood of email, phone calls, faxes, and the like (even putting aside email spam or junk mail). Some people are driven to extreme actions. For example, Don Knuth, one of the most eminent computer scientists, has stopped using email. His attitude is that he does not wish to be distracted from the serious investigations he wishes to carry out. Anyone wishing to contact him has to overcome the extra barrier of using ordinary mail. And that is indeed the beauty of point-to-point communication, that we do have a choice. We are breaking away from the artificial constraints of earlier technologies. A great river of information is flowing at an ever faster pace, and we get to choose what we wish. Based on history, it seems safe to predict that Knuth’s choice will not be typical, and that most people will choose to drink from that river. However, they will be increasingly selective.

For a communication service provider, the task is to make life for the customer as painless as possible. A seamless integration of the Internet into people’s lives should be the goal. One of the most successful companies in this area, at least so far, has been AOL. An AOL press release, dated February 2, 2000 [AOL], claimed 21 million members, as well as usage records, as follows:

Among the AOL service’s other usage records, underscoring that it is becoming more and

more essential to members' everyday lives, are:

...

- ** 110 million e-mails sent daily (50 million a year ago);
- ** 600 million instant messages sent per day (400 million a year ago);
- ** 200 million stock quotes daily (113 million a year ago); and
- ** 5.2 billion Web URLs served a day (2.1 billion a year ago).

It is clear that AOL measures its success by how deeply it is involved in its customers' lives. As we move towards broadband always-on access, it will be possible for carriers to provide even greater services. The question will be how to choose the quality and price combinations to encourage greatest use. Efficiency is likely to be of secondary value.

4. The Internet and other communication services

The Internet is the latest addition to the communication infrastructure that has been developing over the centuries. Most of the discussion in this work is limited to the 19th and 20th centuries. That was the period when the modern industrialized world was born, when the basic power, water, transportation, as well as communications infrastructures were built. For nice, concise accounts of the development of some key infrastructure industries, see [Friedlander1, Friedlander2, Friedlander3, Friedlander4]. For an earlier unified account of the development of U.S. communications in the 19th century, see [John3]. This section provides a brief overview of some of the important similarities and differences between the various communications technologies, especially as they affect the main concern of this work, namely pricing and service differentiation.

Many similarities between the Internet and earlier communications services have been noted in earlier sections, and many more are presented in this one and later ones. The analogies between the Internet and the telegraph are especially strong. Many are presented in the fascinating book of Tom Standage [Standage], which calls the telegraph the "Victorian Internet." However, there are also numerous similarities to other services, as well as common themes running through the combined history of all of them. The most notable for the purposes of this work were the growth in usage and decreasing prices, discussed in the preceding section. However, there are many other similarities that might be useful to consider. Some are just amusing. For example, today's carpal tunnel syndrome was known in the late 19th century as "glass arm" or "telegrapher's cramp" (p. 18 of [Jepsen]). The complaints (less frequent

now than half a dozen years ago) that traffic between European countries was getting routed through the U.S. recall periods when it was faster to send a telegram between London and Paris through New York than directly. Other repeating patterns might have more substantial lessons for us. As an example, most new technologies led to new forms of crime, and certainly to fears of crime. In late 1999 and early 2000, denial of service attacks on the Internet as well as other, real or potential problems, led to concerns that anonymity on the Internet might be harmful. As a result, in early 2000, the French National Assembly was debating a law that requires Web site creators to identify themselves to their hosts. This is strikingly similar to what happened 150 years earlier. The electric telegraph in France was initially limited to government use. This was finally changed in 1850, but with the restriction that official messages would have priority over commercial ones, as well as with restrictions on anonymity.

Similarly, to prevent criminal elements from ‘using the telegraph to hatch their detestable plots with more ease and rapidity’, the law demanded that private dispatches be sent un-coded and signed, and that the user declare his or her identity. This constraint was added by the Assembly which changed the government’s wording (‘any person is allowed to correspond’) to ‘any per person whose identity has been established’.

p. 46 of [Flichy]

The current debates about privacy on the Internet might benefit from consideration of an earlier period where privacy was largely lost for a time. This is illustrated by the following story published in 1909:

Henry Winston, a prosperous young farmer living near Fairfax, Missouri, did not reckon with the rural party line, to which his telephone was connected, when he called up Miss Lorena Simpson and asked her to share his joys and sorrows for life, and for that reason is in quite a predicament. Before Miss Lorena could answer his question, eleven fair damsels, who had heard, at their respective homes, the ring for the Simpson home, quickly stepped to the telephone. Each, knowing full well that “something would be doing,” had quickly answered “Yes,” adding, “You set the date and I’ll be on hand.”

Although Henry knows he is engaged, and that it is not Miss Simpson who accepted, he is a little in doubt as to which one of the eleven answered his question first.

[Anonym2]

While the story was surely meant as a joke, it does shed light on the social environment of that period (as most jokes do). One can use it as an illustration of gender relations then. For the purposes of this

work, though, it is more interesting to note that this joke assumes an almost complete lack of privacy on telephone party lines. That eavesdropping on such lines was endemic, although against both telephone company rules and against generally accepted standards of good behavior, has also been confirmed by more scientific studies [Fischer2]. Yet people voluntarily accepted this loss of privacy in exchange for a lower cost of service. It was universally recognized that single lines were preferable, and the problems of party lines (such as that of Henry Winston above, as well as less serious ones, such as not being able to use a line at one's convenience at any time) did indeed eventually lead to almost every household switching to single lines. Just how pricing and growing affluence influenced the move away from party lines should shed light on people's willingness to pay for privacy. How people coped with the lack of privacy on party lines could also be a fruitful subject for sociological studies that go beyond [Fischer2].

In the remainder of this section we concentrate on issues of subsidies, economies of scale, and the value of a network. A common feature of all communication services has been that prices have usually had only a slight relation to costs. Average prices have, over long periods, decreased in line with the decline in costs. However, the structure of prices has traditionally included cross-subsidies. It is often difficult to tell when such subsidies exist. One could hope that at least the the economics of postal systems would be well understood. These systems are large, relatively slowly changing, and unusually open to public scrutiny. Yet even such understanding is lacking. As an example, one careful study noted that the U.S. Postal Service's "efforts to determine the cost of the various services provided have been unsatisfactory" (p. 51 of [Sorkin]). In a complicated system that has large fixed costs and provides many services that draw on the same basic resources, cost allocations are by necessity largely arbitrary. The arbitrariness of prices is also suggested by the wide variation in pricing and service class offerings for mail services in different countries. Even though essentially all postal systems have been government run for the last century, and universally operate under a mandate to provide a public service that recovers its costs through usage charges, different countries have adopted different approaches to pricing.

In many cases, though (as we will see later), it is clear that there are or were large cross-subsidies. An advantage of a long historical view, such as that of this paper, is that it allows us to avoid consideration of the minute differences in cost allocations that fill millions of pages of regulatory and court filings. Instead, we can concentrate on the large picture, where it is often much easier to see what was happening. Sometimes these cross-subsidies were mandated by the government. That is how the current huge transfers from long distance phone companies to the local ones in the U.S. through high access charges are maintained. In other cases they were imposed by private carriers. It is often claimed

that a competitive market (such as was supposed to be set up in the U.S. by the Telecommunications Reform Act of 1996) would eliminate cross-subsidies. However, the evidence argues strongly against this simplistic view. Even in competitive markets, service and good providers often have strong incentives to discriminate in pricing. (These incentives, and the socially desirable role that such cross-subsidies play, are described briefly in Section 5.) In the airline industry, deregulation led to the rise of complicated pricing schemes that shifted the costs onto business users, to the gain of vacation travelers. A similar phenomenon can be seen historically in the development of the telephone industry. There is a fascinating story, briefly described in Section 14, of the effects of competition between the Bell System and independent phone companies around 1900. Their rivalry led to an increase in price discrimination in favor of residential users, and a greater shifting of the costs onto business customers.

It is worth noting that subsidies are sometimes voluntary. For example, a large fraction (sometimes estimated as high as 40%) of long distance calls in the U.S. are to toll-free 800 numbers. (As a result of their popularity, they have been joined by similarly free 888, ..., numbers.) Commercial enterprises as well as government agencies pay for these links to make it easy to be reached.

Communication services differ greatly in the degree to which they benefit from economies of scale and technological progress. Postal systems certainly did gain from both factors in the 19th century. Growth in volume of traffic and greater population density combined synergistically with the higher speeds and lower costs provided by railroad transportation to yield much better and lower cost service. (However, as often happens with new technologies, railroad transport initially tended to raise costs.) On the other hand, advances have been much slower in recent decades. The main problem is that mail is still extremely labor intensive. The U.S. Postal Service (USPS) has 770,000 regular employees. The entire U.S. telephone industry has about 30% more workers, yet has about four times the revenues (see tables 2.1, 3.1, and 3.4). This is because about 80% of USPS costs are in wages and benefits for personnel, and technology is only a small fraction of the budget. Over the last two decades, the volume of material delivered by USPS has doubled, while employment has gone up about 30%. Hence productivity has been growing at only 2% per year, far below that of high-tech industries. Automation inside the system and greater use of presorting by large mailers help, but the final delivery still requires extensive and expensive human labor. The marginal cost of delivering a letter is substantial, by some accounts over 60% of total cost. (As with other services, it is hard to define marginal costs properly, and a single letter traveling in a slack period could arguably be claimed to have zero marginal cost. However, many temporary helpers are hired for each Christmas peak mailing season, so it is reasonable to accept that marginal costs are high.)

The telegraph did benefit greatly in its early decades from both economies of scale and technological progress, and this is reflected in the prices in tables 13.1 and 13.2. However, there were limits here also, since telegrams had to be handled by human agents at both ends. (This included delivery to sometimes distant locations, although as the 20th century progressed, and the telephone became widespread, physical delivery of the telegram was increasingly replaced by reading it over the telephone.) As a result, prices were always high. In the end, when traffic volumes started declining, a vicious spiral of price rises and decreases in usage led to the demise of this service that had once been key in knitting the world together.

The history of the telephone shows a complicated mixture of cost trends. There have been very rapid technological improvements. In recent decades, those have led to rapid decreases in costs of long distance communication, and moderate improvements in local links. There are also substantial economies of scale in both local and long distance telephony. However, these economies did not always apply. During the first few decades of its existence, the phone industry suffered from the opposite trend, namely substantial diseconomies of scale. (For more detailed coverage of this subject, see [Friedlander2, Gabel1, Gabel2, GabelW, Mueller] and the references there, as well as Section 14 of this paper.) There were some economies of scale in deployment, and the learning curve reduced cost of equipment as shipments rose. However, there were diseconomies of scale in local operations. Switching was done manually, so an operator was involved in every call. Further, as the size of an exchange grew, calls increasingly required cooperation of several operators. These diseconomies of scale led to pressure for rates that increased with the size of the local serving area, as well as for metered local rates. (Section 14 discusses this at greater length.) The economic logic against flat rates in such an environment was overwhelming, and their survival in the U.S., discussed in Section 14, is worth considering carefully. In today's voice phone system, the case for flat rates, both for local and long distance calls, is much better. Technology has reduced the marginal costs of carrying traffic, and most of the cost is in maintenance of the local connection.

The economics of the Internet are complicated, and hard to summarize in a few sentences. (See [MacKieMV1, Odlyzko7] for overviews.) For one thing, they are inextricably intertwined with those of the phone system. A major reason the Internet developed as quickly as it did is that it could take advantage of the infrastructure of telephony. By now, though, the bandwidth of the Internet and of private line data networks is far larger than that of the voice network, and so it will be developments of the Internet that will matter the most for telecommunications in the future.

The key point in the studies of [MacKieMV1, Odlyzko7] is that the costs of running the Internet

are largely independent of traffic. Thus marginal costs are zero up to the point where congestion occurs and forces addition of new capacity. Also, as is detailed in [Odlyzko7], most of the costs of the Internet are at the edges of the network. This is a different environment from that of the postal system, the telegraph, or early voice telephony. It makes the case for flat rates more convincing than for the other services.

Some communication technologies have disappointed their proponents. ISDN, early teletext systems, and (at least so far) wireless data are prominent examples. On the other hand, the main communication services have usually surprised their managers by growing more rapidly than expected. There have always been visionaries who correctly forecasted rosy futures. (The book [deSolaP2] has some nice examples.) However, the people in charge of running operations have tended to be too conservative. This was true of the British postal officials who scoffed at Rowland Hill's predictions about the increased volume of mail that would result from lower rates. (See Section 12.) It was also true of the U.S. postal service, which underestimated the demand for personal correspondence in the early 19th century [John1]. It was certainly true of phone companies.

In the very earliest days, Vail had expected that the highest level of development would be one telephone per 100 people... In 1905, a Bell estimate assumed that twenty telephones per 100 Americans was the saturation point and even that "may appear beyond reason." ... One reads in Bell documents of the late 1920s of concern that the automobile and other new technologies were far outstripping telephone diffusion. Yet, even then, there seemed to be no assumption that the telephone would reach the near universality in American homes of, say, electricity or the radio.

p. 57 of [Fischer1]

B. Bjurel, the head of the Swedish telephone company, said in 1976 that

Many telephone administrations have made great mistakes in their assessment of the future development of telecommunications. They have been altogether too pessimistic concerning the possibilities of expansion. Both as regards number of subscribers, telephones, and traffic, certain saturation values were assumed at an early stage. These values have constantly proved to be too low.

(quoted on p. 85 of [Kraepelien])

In recent times, the potential for cell phone penetration has often been underestimated. Most notoriously, the McKinsey consulting firm predicted in the mid-1980s that by the year 2000, there would be

fewer than a million cell phones in use in the U.S..

Historically, there has been a pronounced tendency to underestimate the potential of communications services to facilitate social interactions. High postal rates discouraged such correspondence at the beginning, and this led to the perception that there was little demand for it. For example, “as late as 1830, Tennessee congressman John Bell could term the high letter fees ‘popular,’ since they helped to prevent the postal system from running into debt” (p. 156 of [John1]). There was similar skepticism in the U.K. about Rowland Hill’s proposed Penny Postage (discussed at greater length in Section 12). The telephone also carried primarily commercial traffic in its first few decades. It was only in the 1920s that it discovered that sociability was the key to much greater penetration and increased revenues [deSolaP1, Fischer1, Fischer2].

Finally, the Internet has been constantly surprising with its rate of growth. Even while it was just a research network, there were concerns that its growth rate was about to slow down. There are still persistent doubters. For example, Noll predicted in 1991 that voice traffic would always be far larger than data traffic (pp. 171-175 of [Noll1]). Since that time, the volume of data traffic has surged far beyond his original prediction. (The main reason for this development is that he did not foresee the arrival of graphics-rich content, such as Web pages, and was considering only text.) Still, he continues predicting that data traffic will not exceed voice traffic unless multimedia begins dominating the Internet [Noll2]. While he is apparently right that there is still more voice traffic than data traffic (cf. [CoffmanO1]), his estimates for data are far too conservative. There is more traffic than he estimates. Further, the Internet is continuing to grow, even without multimedia. One hears comments such as “[n]o one anticipated the volume of requests for [consumer] high-speed connections to the Internet,” coming from a local carrier that is offering ADSL connections.

While there are some Internet skeptics, such as Noll, there are also many wild enthusiasts who talk of Internet traffic doubling every three months. While such growth rates (approximately 1,000% per year) apparently did hold in 1995 and 1996, since that time the rate has slowed to about 100% per year, when traffic is measured in bytes [CoffmanO1, CoffmanO2]. While that is not a “doubling every three months,” it is an extremely rapid growth rate, much more rapid than has been seen over extended periods in any other technology.

The doubling of Internet traffic every year is not guaranteed to continue. However, there is no reason to think it cannot. Technical progress in transmission and switching appears sufficient to sustain growth at these rates for a decade at least.

Where will the increased Internet traffic come from? This question is considered in more detail in

[Coffman02]. The basic answer, though, is that we don't know for sure. Packetized voice, movies, as well as a variety of ecommerce transactions are all likely to play a role. The key point is that most of the growth will have to come from more intensive use of the Internet by individual users. There will certainly be growth in users. However, that is likely to slow down. The ranks of French Internet users, shown in Table 9.2, are still growing at about 100% per year, but that is only because they started from a low base. In the U.S., growth in residential subscriber ranks in 1999 was down to around 40%. When more than half of the population is on the Internet, you cannot double the participation rate! However, you can increase their usage, essentially without bound. For example, the spurt in Internet traffic in 1995-6 appears to have been the result of two closely related factors. Millions of users of traditional online services, such as CompuServe, Prodigy, and AOL, discovered the Internet. In addition, the text-oriented content was replaced by graphics-rich Web pages. As a result, there were more users, they were spending more time online per person, and while they were online, the data transfer rates increased.

Table 4.1. Growth in Internet traffic.
Volume of data received by the University of Waterloo during each month.

month	MB/day
1993-03	453
1993-07	541
1993-10	648
1994-04	977
1994-08	1,377
1994-11	2,098
1995-04	2,285
1995-07	3,229
1995-11	6,000
1996-03	7,229
1996-07	7,823
1996-11	10,962
1997-03	11,463
1997-07	12,087
1997-11	24,621
1998-03	24,676
1998-07	22,572
1998-11	39,808
1999-03	46,720
1999-07	43,862
1999-11	68,582
2000-03	74,304

A particularly interesting example of data traffic growth is presented in Table 4.1. It shows data flow from the public Internet to the University of Waterloo in Canada. Student ranks have not increased over the last decade, yet traffic has continued to grow. The rate of increase has slowed down to about 70% from 100% per year, but that is still fast growth. This growth represents more intensive use by the same population, the kind of growth that is bound to characterize the Internet in the future. Thus one relevant measure of success for Internet service providers will be how much data their customers send and receive, and not just how many customers they have.

For content delivery, the value of a network is generally regarded as proportional to the number of users. That means that the value of half a dozen separate networks is equal to that of a single network. Hence there is little to be gained from general interconnections. In particular, if the Internet were to be primarily a content delivery network, the economic forces pushing for balkanization might prevail. There would still be advantages to some standardization. The incompatible standards of CompuServe, Prodigy, GENIE, Microsoft Network and AOL in the early 1990s increased costs to users and service providers and greatly impeded the penetration of online services. However, one could use the same standards and not have interconnections, or at least not easy interconnections. There are historical precedents for territorially overlapping networks that do not interconnect, even in point-to-point communications. Especially interesting was the period around 1900 when there was intense competition in the U.S. between the Bell System and independent phone companies, often operating in the same territories. It often happened that nearby houses were connected to different carriers, and could not communicate by telephone. Some residential users, and a larger fraction of businesses, coped with this obstacle through the inconvenient device of having two telephone lines, one from each competing company. (There are more details about this development in Section 14.) At one stage the independents had about as many subscribers as the Bell System. Although the arrangement appears inherently inefficient, some scholars, led by Milton Mueller [Mueller], argue that it facilitated the spread of the telephone by stimulating competition. Hence one could argue that in the early days of rapidly growing service, separate networks might be beneficial in the long run through this same process.

For point-to-point communications, the commonly accepted view was formulated by Bob Metcalfe around 1980, and is widely known as “Metcalfe’s Law:” “The value of a network is proportional to the square of the number of users.” The argument behind this “law” is that each member of a network derives value from it proportional to the number of other members that can be reached. Hence adding up the valuations of all members, we obtain a total value that is quadratic in the size of the network. If this “law” were indeed valid, there would be no need to worry about general connectivity. Combining

two equal size network would double the total value, and so the two carriers would have an irresistible economic incentive to link up, either through merger or some standard interconnection. Public policy concerns could then be limited to questions about monopoly power. Connectivity would take care of itself.

Unfortunately, the valuation of a network provided by Metcalfe's Law is not very accurate in many cases. If population doubles, and the technology stays the same, it is hard to imagine the value of the network quadrupling. (If it did, a further doubling of population would increase the value of the network 16-fold over the original one, and eventually the value of the network would be larger than that of the whole economy.) We also have historical evidence that Metcalfe's Law sometimes fails. Consider the early 1900s in the U.S., when the independent phone companies were increasingly interconnecting with the Bell System. If Metcalfe's Law were valid for those networks, huge extra value should have been created. This value should have been reflected in either much higher prices for phone service, in greater profits for the phone companies, or in many more users joining the system. While phone industry industry and profits did increase, they did so at moderate rates. Further, prices generally declined, and the growth in ranks of phone subscribers actually slowed down. Thus history throws doubts on the validity of Metcalfe's Law, at least in its literal sense. However, this law does reflect a fundamental truth, that the value of a point-to-point communication network increases faster than just in proportion to the number of users.

Metcalfe's Law was formulated relatively recently (as were other, slightly earlier, quantitative estimates in the economics literature). However, the basic insight about the value of a point-to-point communication network growing faster than linearly, is old. It guided the development of the U.S. postal system (cf. [John1]). It was also a key principle in the development of the phone system. For example, a century ago, it motivated the Bell System in setting its price structure:

AT&T also favored [local measured service pricing] because they believed that it would increase network membership. The president of AT&T, Frederick Fish, believed that customers valued access and that charging a low fee for network membership would maximize the number of subscribers. According to Fish, the number of users was an important determinant of the value of telephony to individual subscribers. His desire to maximize network connections led the firm to adopt a pricing structure in which prices to residential customers were actually set below the marginal cost of service in order to encourage subscriptions. These losses were made up through increased charges to business customers.

[NixG]

Thus Bell System policies were inspired by the same insight that leads to Metcalfe's Law. However, this is not the full story. The value of a network depends not just on the number of users, but also on their potential interactions. Although people value the ability to contact anyone on Earth, in practice most interactions are with a small and slowly changing group. (See, for example, Chapter 10 of [deSolaP1].) Commercial translation services have not fulfilled some early enthusiastic predictions because few people have the need to talk to others who do not speak a common language. The more interactions are localized, the weaker the case for universal interconnections. In the early 1900s, the overwhelming majority of phone calls were local. Many of the independent carriers did not interconnect with others, and thus were offering exclusively local service. Even the Bell System carried few long distance calls. In addition, phone penetration was low. As ranks of subscribers grew, demands for interconnection intensified, and eventually did lead to the current system where any phone can call any other. (See [Mueller] for details of this interesting transition.) Sections 3 and 21 show that we have not yet achieved the famed "death of distance," but the general trend is towards interacting, whether for social or business purposes, over greater distances. This makes it unlikely that any carriers could form a separate network. Corporate customers (who currently generate most of the traffic) certainly would not tolerate it. It is even questionable whether residential customers would accept it. A particularly interesting case to watch is that of AOL and its instant messaging or "chat" system. AOL has been adamant about not allowing other ISPs to connect to it. However, in the past, AOL has in the end (and even without government pressure) yielded to customer pressure to exchange email with other carriers, and then to offer flat rates. It seems likely that at some point AOL will also have to capitulate on instant messaging and allow interconnection.

Metcalfe's Law was shown above not to be accurate in describing the value of a network. The value of a network grows more slowly than quadratically in the number of users. Further, there is another factor determining the value of a network that Metcalfe's Law misses. This "law" concentrates on the number of users in a network, and not on the intensity of their participation. That may be appropriate for some contexts, but, as is discussed above, not for the Internet. It could even be that many decision makers in the past were misled by following the same idea, that maximization of the number of users was the goal. That was certainly the belief of AT&T officials a century ago, as shown in the quote above about Frederick Fish, the president of AT&T. This is important in the discussion of pricing structures, since flat rates seem to be an invaluable tool for maximizing usage. As Nix and Gabel explain,

Fish's belief that value was created largely through the number of customers connected to the network, rather than the number of calls, made it difficult for him to recognize the utility of flat-rate service as a pricing policy to preempt entry. But there were other possible ways to conceptualize the process of network formation. AT&T officials could have conceived the value of network participation as based largely on the number of connections, rather than the number of members connected to the network. Maximum use of the network for a given number of subscribers would have been achieved under flat-rate service.

[NixG]

In the end, against the judgement of phone company officials (both in the Bell System and among the independent phone companies), flat rate did survive, apparently largely a byproduct of the competition. This then led to increased usage, and higher revenues for the phone companies in the U.S. than in most of the rest of the world. It also promoted the rapid spread of the Internet.

5. The effectiveness and utility of pricing

For any service to be viable in the long run, it has to recover its costs. Even the supposedly "free" Internet is nothing of the sort, as it is largely paid for through monthly fees by individual users or their employers or universities.

How costs are recovered from users is a matter of pricing. This section is a brief overview of the main purposes that pricing can serve, aside from the fundamental one of cost recovery. For example, the choices that users make among different quality and price combinations provide valuable information for the service provider about what the customers wish and are willing to pay for.

In most of this paper, flat rate pricing is presented as the ultimate of simplicity. Still, it is useful to note that under some circumstances charging a fixed rate for each byte transmitted would be simpler from some points of view. For example, many of the flat rates, for example for dial modem Internet access, and even more for cable modem or DSL links, are not all that simple when one reads all the fine print. Invariably there are prohibitions on running Web servers, reselling the service, and various other activities. These are understandable. For example, a cable TV company can afford to provide broadband access to the Internet at a low fee, but only if the overwhelming majority of the customers are light users. It cannot afford to have any one signing up for the \$40 per month service and then using that connection to provide Internet access to a whole apartment complex. With usage-sensitive pricing, such arbitrage could become a non-issue, and the fine print in the contract could be dispensed

with. Therefore the conditions of service could be simpler with usage sensitive rates. However, from the point of view of the overwhelming majority of users, flat rate pricing is far simpler. They do not have to think about details of their behavior online, as it typically falls within the scope of the contract.

There is a wide spectrum of options for pricing communications services. To illustrate, let us consider the regular postal service. At one extreme one could have a scheme in which the mail carrier would collect letters, and for each one would bargain with the sender, taking into account the destination, the length of the letter, the likely interest the sender has in getting the letter delivered, how heavy the carrier's mailbag is, and so on. Moving slightly away from this extreme towards greater simplicity would be a fixed schedule of charges, covering each possible source-destination pair of houses. Such a schedule might be based on some allocation of postal costs, and would depend on how close to the nearest post offices the sender and receiver live, how much other traffic arrives at or leaves their houses, and how much mail their neighbors originate. A simpler scheme yet is one in which the charges depend just on the distance between the origin and destination. (A more complicated version would take use the distance along the actual physical route taken by a letter, a simpler one would depend on just the air distance.) Simpler yet would be a scheme that has a fixed charge, independent of distance. Finally, the simplest system of all would have no charges at all, and would provide free service, paid for by general taxation.

In addition to the range of options sketched above, there are several others that can be added, almost orthogonal to that one. For example, one can have, for most of the options listed above, time-of-day pricing, or quality differentiation with respect to speed of delivery.

Most of the feasible price and quality combinations have been tried in the past. For example, before postal systems became widespread, getting a letter delivered required finding somebody traveling in the right direction and negotiating for delivery. On the other hand, at least for some time in the mid-1850s, Iceland had mail delivery paid for out of general taxes [Miles]. In the early 19th century in the U.S., apparently there were also some calls to make the mail a free public service (p. 62 of [Baratz]). These proposals were never considered seriously for all mail, but free delivery of newspapers was debated repeatedly in Congress, and was practiced in a limited way some of the time. (This was an instance of policy makers promoting content at the expense of point-to-point communication. In the U.K. in the early 19th century, newspapers were carried by the postal service at no charge, but for other reasons. See Section 12 for more details.) Historically, from the early 18th century until today, the postal system has moved from charges depending in complicated ways on source, destination, and route, to a uniform fee per piece, up to some weight or size limit that few items are likely to reach. However, there is

service quality differentiation, with several classes of mail.

On the Internet, many pricing proposals call for complicated schemes with charges varying from packet to packet. Such schemes can be proved to have various optimality properties, whether the basic philosophy is to have charges closely related to costs, or to maximize revenue. (For one such proposal, employing a Vickrey auction, which has been tremendously influential in stimulating thinking about Internet pricing, see the paper of Jeff MacKie-Mason and Hal Varian [MacKieMV1].) On the other hand, most charges today are through flat rates. There is little advocacy for public funding of the Internet, other than for research networks, although a few calls are still heard occasionally for public funding of peering points, to avoid the sticky interconnection problems. The general attitude is that communications users should fund the services they consume. A competitive service environment certainly requires that. The question is, how closely should one attempt to link the fee that a customer pays to either the costs that customer's usage generates, or to the value that customer derives from the service?

Historically, the relation between cost of providing a communication service and the price charged for it has been slight, as we will see repeatedly in later section. Cross-subsidies and price discrimination have been common.

I first discuss the desirability and frequent necessity of price discrimination. In our society, the word discrimination has a strongly negative connotation. However, price discrimination is a common business tactic that is growing in importance, and serves a useful social purpose. Consider a simple example. (For detailed discussion of the different types of price discrimination, see [Varian2, Varian3, Varian4].) Suppose that Alice and Bob run two businesses near to each other at the end of a long road, and both are interested in obtaining broadband access to the Internet through a fiber cable. Suppose that Alice is willing to pay \$10,000 in installation charges for such a link, while Bob is willing to pay \$20,000. Suppose also that the cost of bringing the cable to their area is \$25,000. If a carrier has to charge the same price to Alice and Bob, neither will be able to get the service. Any charge of \$10,000 or less per house would lead to both Alice and Bob signing up, but would not recover the \$25,000 cost. A price of over \$10,000 per house would lead to only Bob signing up, and he would do so only as long as the price did not exceed \$20,000, which again would not allow a carrier to provide the service. On the other hand, if the carrier can discriminate, and charge Alice \$9,000 and Bob \$19,000, say, the transaction will take place. In the conventional economic view, this outcome is to be applauded, since both Alice and Bob will get the service they want for less than they value it at, and the carrier will make a profit.

In practice, any seller that wants to price discriminate has the difficult task of finding out the customers' willingness to pay and getting them to accept the discrimination. A common way to solve this problem is to impose artificial restrictions that will induce the customers to sort themselves out. In the example above, the carrier might offer a 10 Mbps service for a fee of \$9,000 and a 50 Mbps connection for \$19,000, with the hope that Alice will purchase the first and Bob the second. With a large number of potential customers, it is not necessary to know individual willingness to pay. Estimates of the distribution of valuations often suffice to set a tariff that lead customers with different valuations to purchase different services. Such sorting out of users is the main purpose of the advance purchase and Saturday night stayover restrictions on inexpensive airline tickets. The assumption that carriers make is that business travelers will often need to make last-minute changes in their schedules, and also will be unwilling to extend their time away from home over a weekend. This does work much of the time. It allows airlines to offer inexpensive fares to the vacation travelers who can adjust their schedules to fit the restrictions, and who might not travel at full fares. The result is that often there are two passengers sitting next to each other on the same flight, but one of them has paid more than five times as much as the other one. The system does serve to produce high utilization, decreasing fares, and also availability of seats on most flights even at the last minute (although at high prices).

The airline yield management system does work, but it is generally disliked by the public. People find the bewildering array of constantly changing prices and conditions unpleasant, exasperating, and unfair. One way to make price discrimination more palatable to the public is through quality differentiation. The airlines practice that also by providing several classes. First and business classes get comfortable seats and good meals, while coach passengers are squeezed into tiny spaces and receive minimal meal service. This is not a new tactic, as it was observed in operation a century and a half ago:

It is not because of the few thousand francs which would have to be spent to put a roof over the third-class carriages or to upholster the third-class seats that some company or other has open carriages with wooden benches. What the company is trying to do is to prevent the passengers who can pay the second class fare from traveling third class; it hits the poor, not because it wants to hurt them, but to frighten the rich. And it is again for the same reason that the companies, having proved almost cruel to the third-class passengers and mean to the second-class ones, become lavish in dealing with first-class passengers. Having refused the poor what is necessary, they give the rich what is superfluous.

Emile Dupuit, 1849

(as quoted in [Varian5])

It appears that this type of price discrimination through versioning, in which there is a perceptible difference in quality, is more palatable to the public than that of airline yield management, where seats in the same class are sold at wildly varying prices.

Some types of price discrimination are so traditional that they are accepted without comment. For example, postcards were introduced first in Austria in 1869, and quickly copied by other postal systems as a result of their popularity. The price for mailing them is considerably lower than it is for first class letters. However, they cost just about as much to handle. The role of the lower prices for postcards is to segment the market, with the correspondence that is short and does not require privacy paying less. In a similar vein, books also offer a case of rampant price discrimination.

Although people do not think of [the book trade] this way, current practices involve charging different prices to different users, and thus maximizing revenues. A novel is typically published in hard cover first, with the aim of extracting high prices from those willing to pay more to read it right away. Once that market is fully exploited, a somewhat cheaper trade paperback edition is made available, to collect revenue from those not willing to pay for the hardbound copy. Finally, a regular paperback edition is published at an even lower price. The used book market develops in parallel, for those willing to read books marked up by previous owners, and so on.

[Odlyzko2]

The incentive to use quality differentiation to price discriminate is even leading to increasing application of the “damaged goods” approach, in which the seller incurs increased costs to provide goods that sell for lower prices. This phenomenon is explored in detail in [DeneckereM]. One example cited there is of two IBM laser printers of around 1990, in which the slower and less expensive one differed from its higher priced almost-twin just in having some extra circuitry that slowed it down. Another example is of Federal Express service, which has next day morning delivery, and a less expensive next day afternoon delivery. Packages sent via regular overnight rates are not delivered in the morning, even if they arrive in time for morning delivery, and even if it requires an extra truck trip to make the afternoon delivery. In the electric power industry, it is widely reported that purchasers of “interruptible supply” contracts (in which the utility reserves the right to stop the current flow when overload conditions occur) occasionally have their supplies cut off. This serves to ensure that users who do require a continuous supply do not gamble on interruptions being rare.

The “damaged goods” method violates many people’s sense of fairness, and incurs additional costs. However, it does help to price discriminate among users. It might be required on the Internet (just as it is in interruptible electric power supplies) if quality differentiated transmission, such as that of the currently popular diff-serv model, is deployed on a wide scale. The problem is that the natural variation in traffic on the Internet means that for much of the day, including sometimes moments during the peak traffic periods, one might be able to get high quality transmission for traffic sent at the lowest priority. This might mean that the traffic sent at higher priorities might be too small a fraction of the total to make differentiated services economical. It would likely be necessary to introduce artificial degradation of service quality for the low priority traffic (as was proposed in [Odlyzko5]), say through artificial delays or packet losses. For example, the lowest cost channel might have enough degradation introduced into it to make either video or voice transmission unacceptable. The next higher priced channel might have enough degradation to allow video, but not high quality voice. Finally, the highest priced channel might have quality high enough for telephony.

Aside from price discrimination, there is another, related, reason for using pricing schemes more complicated than the flat rate one, namely to balance the load on the network. The problem of electronic communication networks is that most of their costs are independent of traffic. (That is not the case for postal systems, and is worth keeping in mind when reading about the different pricing and quality differentiation policies.) Measures that even out the load can therefore reduce costs of the system. Such measures could be short-term (even on an individual packet basis, as has been proposed for the Internet by MacKie-Mason and Varian in their famous “smart markets” scheme), although that has not been tried yet. These measures can also take the form of simpler time-of-day pricing schemes, which have been widely used, and are being extended to road tolls, for example.

It is often hard to disentangle load shifting and price discrimination motives when time-of-day pricing is present. However, often it was clearly the desire to charge according to value rather than the desire to even out the load on the network that was dominant. Examples, presented later in the paper, include the initial introduction of evening discounts by the Bell System.

Price discrimination is sometimes hard to practice. It is necessary to obtain information about customers’ willingness to pay, and they are naturally reluctant to provide it directly. Furthermore, price discrimination is often resented, and is sometimes illegal. As one example, many cities in the U.S. have passed laws making it illegal to charge more for dry cleaning a woman’s shirt than for a man’s.

Price discrimination can be made more palatable by careful disguise. For example, the selective private colleges in the U.S. are engaged in an extreme form of price discrimination. They charge very

high tuition, and at the same time have a needs-blind policy, in which any admitted student is guaranteed a mix of scholarships, loans, and part-time jobs that covers the gap between what the college expects the parents to be able to pay, and the school fees. This is in effect a simple case of charging according to the ability to pay, but is disguised through the ruse of scholarship support. Coca Cola's plans to vary the price of soft drinks would surely have been received better had they been presented as aimed at stimulating sales when it is cold by lowering prices. (The discounts that senior citizens are often offered are a blatant form of price discrimination that would undoubtedly attract more negative reaction were it to be presented in the form of a premium charged to everyone else.) Human reactions often do not conform to the standard economic models of marginal utility, and have to be taken into account.

So far we have discussed arguments for price discrimination. On the other hand, there are many cases where there incentives for such discrimination clearly exist, but it is not practiced. For example, voice services bring in huge revenues. On the other hand, they require low bandwidth. Hence on a broadband network, if transmission of video is to be affordable, and charges are to be based on bytes transmitted, voice should be essentially free. Yet voice services is still what provides the overwhelming bulk of revenues to the communications industry. Hence the argument has often been made that voice transmission should be charged at higher rates than video. This type of price discrimination is appealing on public policy grounds. However, the same argument would argue for high charges for email. After all, email is still the "killer app" of the Internet, perhaps even more than the Web, and the bandwidth it requires is modest compared to Web surfing. Moreover, special email charges would (as many people have suggested) help control spam. Yet, historically, charges for email, which used to be common in online services, have been phased out. (Not only that, but we have a multiplicity of services offering free email accounts.) And indeed, with scores of email messages per day for most "wired" people, who would want to deal with the extra complexity of having to decide for each one whether it is worth paying for it?

The general trends in communications, as is documented in later sections, has been towards decreasing price discrimination, and towards simplicity. On the other hand, there is also a tendency towards complicated pricing, such as yield management strategies, auctions, and the like. This is evidenced not only by the airlines, but also by the successes of eBay and priceline.com. My conclusion is that the conventional economic arguments for quality and price differentiation are all valid. However, they conflict with some basic human desires. For inexpensive and frequently used goods and services such desires are strong enough to overcome the economic push for optimization.

It is hard to make categorical statement about human preferences, since people exhibit contradic-

tory traits. They often simultaneously increase and decrease their risks, as when they buy both lottery tickets and insurance. However, generally speaking, people are risk averse, especially in the face of large potential losses. That is why the insurance industry is vastly larger than the gambling industry. In general, people are also averse to varying prices, and are more willing to accept variations in quality than in price, as is illustrated by many examples in [Odlyzko5]. That paper also discussed such economic puzzles as the laws against scalping. Such laws are a puzzle for two reasons. One reason is that according to standard economic doctrine, scalpers provide a useful function, by moving desirable tickets into the hands of those who value them most highly. Thus their business should be encouraged, and not driven underground. Another reason such laws are a puzzle is that from the conventional economic perspective it is hard to explain why they should be necessary at all, no matter how one views scalpers. Why is there room in the economy for scalpers? Why don't the owners of sports arenas and concert halls adjust prices to clear the market in the first place? The most satisfactory explanation seems to be that both sports arena operators and lawmakers are reacting to human preferences for simple, predictable, and fair pricing. Other examples of similar seemingly irrational behavior are cited in [Odlyzko5]. In most cases, though, such examples serve as illustrations only, and do not provide hard estimates. In general, it appears likely that appearing to follow general notions of fairness will be increasingly important in ecommerce ([Odlyzko2, Zajac]).

6. The role of distance dependence in pricing

The most frequently used communication service, and the one that attracts the most spending, is the phone. Charges for phone use have traditionally depended on the distance of the connection. Local calls were usually the least expensive, and prices of long distance calls rose steeply with distance. The one common communication service whose tariffs did not vary with the distance was the mail. (Even there, foreign letters invariably carried higher fees than domestic ones.) When charges were distance-independent, this was often a cause for comment, and was often referred to as a case of "postalized rates." In particular, the Internet has often been pointed out as unusual in ignoring distance. Occasionally worries have been expressed about the harmful effects of this policy, which ignores the costs of long distance transport. Of special concern have been the expensive trans-oceanic links. In some cases (such as that of the British academic JANET network, discussed in more detail in Section 19) charges have been imposed on member institutions for traffic coming from the U.S., but not from Europe.

While costs of long distance transport are a concern to many Internet experts, the historical trend in other communication services has been to increasingly ignore them. A key role in this development

has been played by technology, which has lowered transport costs. However, this paper avoids detailed consideration of internal costs of service providers, and considers primarily the prices consumers have been faced with. Originally, even postal fees were not “postalized.” They became distance-independent only in the 19th century, starting with the famous British postal reform of 1840, described in Section 12. Since that time, many communication services have traveled either partially or all the way towards abandoning distance sensitivity. This trend in prices is illustrated in tables 12.3, 13.1, 13.2, 14.1, and 14.2. Further decreases in distance dependence are occurring right now. As one example, Telia, the dominant Swedish phone company, announced in February 2000 that all phone calls throughout Sweden would be charged at one of two rates, depending only on time of day, but not on distance. Many more examples will be presented throughout this paper. (There is also an interesting example of the reverse tendency, namely of FedEx introducing distance dependence to its pricing. This example does not contradict the basic thesis of this work, since it is associated with an expensive and infrequently used service. See Section 12 for more detailed discussion.)

A simple argument says that if prices are distance-insensitive, then the basic network costs (of switching and transmission) must be a small part of the price that is charged. But if those costs are low, then the costs of tolerating various inefficiencies (such as carrying low-value communications) are also likely to be low. That argues for a single grade of uniformly high quality service, and simple pricing.

An even more convincing case for simple pricing and simple service offerings can be developed based on a detailed historical study of the distance sensitivity of rates. Frequently such distance sensitivity was far greater than was justified by cost differences. It was a form of price discrimination. This practice reached an extreme form in the U.S. phone industry in the 1970s and 1980s, when long distance revenues were used to subsidize local service.

The precise cost allocation between local and long distance services is an inherently complex subject. The millions of pages of scholarly, regulatory, and legal papers on this problem testify to the impossibility of a definitive solution to this problem. However, the advantage of taking a very high level view is that it is possible to discern general trends without getting bogged down in details. It appears that for a long time, prices for long distance and local service in the U.S. were set to approximately recover the costs of the respective services. Starting in the 1950s, technology led to a rapid decrease in costs of long distance calling. Regulators and the Bell System used only some of these cost savings to reduce long distance rates, and allocated an increasing proportion of the savings to support basic local connectivity. For a slightly more detailed, but still short, account of this development, see [Mueller]. In particular, the graph on p. 161 of [Mueller] provides a very striking illustration of the

growth of the subsidy. Subsidies of that size are incompatible with competition, and so even the slight deregulation in the 1970s led to the rise of competitive long distance carriers and the breakup of the Bell System. A strong residue of this system of subsidies from long distance to local calling remains in the shape of high access charges paid by long distance carriers to local ones. These access charges were mandated by the government. While they are being reduced, until recently they accounted for more than a third of the price of a long distance call. The result is that the U.S. has a bizarre quilt of phone rates. For residential users, local calls are typically free. On the other hand, toll calls to towns outside the immediate vicinity but within the area served by the local carrier are charged according to distance, and often are priced far higher than a call across the continent or even to England that is carried by a long distance company.

Although the extent of cross-subsidies within the Bell System was extreme, charging far more than justified by costs for long distance messages has been practiced by many carriers. Private companies, whether they had monopolies or were in competitive markets, have done this just as government monopolies have. Many examples will be discussed in later sections. It was certainly true of the postal systems of the U.S. and Britain in the early 19th century, before postal rates were “postalized.” It was true of the Swedish phone systems pricing in the 1980s [Mitchell1].

Why would carriers impose charges that increased more steeply with distance than costs? The basic reason is the perception that long distance transmission is more valuable. In U.S. voice telephony, regulators argued that expensive long distance calls were a luxury for the rich, and so could in effect be taxed to provide basic service for the general population. More generally, such practices are a form of value-based pricing. This role has been recognized for a long time. For example, in the discussions preceding the British postal reform of 1840, which did introduce distance-insensitive prices, we find the following claim:

Distance is the chief ingredient—indeed, we might say, the very *sine qua non* and essence—of any and every system of post-office charge... If it were not for distance, there would be no post offices at all. The intrinsic value of the conveyance of a letter... is exactly equal to the time, trouble and expense which is saved to the correspondence—of which the best, if not the only measure is distance; and as the difficulty of private conveyance increases, so much increase, in gradations proportional to the distances, the value of the conveyance. The gods must annihilate both time and space before a uniform rate of postage can be reasonable or just.

J. W. Crocker, *Quarterly Review*, 1839
(as quoted in [Gregory])

Well, gods did annihilate both time and space. The uniform Penny Post was introduced in 1840, and has historically been regarded as a great success.

One cannot argue that all or even most instances of price dependence in pricing of communication services are forms of price discrimination. There have often been substantial cost differences to justify substantial price differences. For example, the first transcontinental calls from New York City to San Francisco, offered in 1915, took 23 minutes to set up, as it required many operators and expensive novel equipment. Even in 1925, the call setup time was over 7 minutes. When transatlantic phone calls were offered in 1927, there was only a single circuit, and a call took hours to set up, which helps explain the high costs seen in Table 14.1. Similar cost barriers to distance-insensitive pricing existed early in the history of the telegraph. Long distance telegrams often had to be retransmitted at intermediate points. Even on a single link, the speed of transmission decreased with distance.

The wide variation in costs of transmitting messages can be used to justify non-uniform prices. However, it is interesting to observe that many of the price schedules that were not uniform depended on a simple formula that took only distance into consideration. (See tables 12.2 and 12.3, for example.) Yet if a carrier were to do what is theoretically socially optimal, namely have prices close to the marginal costs or to willingness to pay, it would have much more complicated price schedules. Costs are only weakly related to the distance a message travels. This will be discussed with some precise figures in Section 23 for the Internet. Costs do vary dramatically. Still, the main determinants of costs tend to be factors such as particular layout of a network, and the traffic on individual links. This appears to have been pointed out in a public forum for the first time by Rowland Hill in 1837, in discussions of pricing in the British postal system. (There are extensive discussions of Hill and the famous mail reform of 1840 in Section 12.) It also appears to have been well known for the U.S. postal system in the first half of the 19th century. At that time transport costs were low along the Eastern seaboard, especially in the Northeast, and high along the frontier. A true cost-based pricing scheme would therefore have separate prices for every source-destination pair. (Such pricing apparently was in effect in the postal service in the British colonies before the American Revolution.) A value-based pricing plan would also surely not depend in a simple way just on distance. And indeed, when we examine the competitive spot markets for long distance bandwidth that are emerging (see Section 23 for details), we find that the old tariffs based on distance are irrelevant, and prices vary from place to place and from day to day. (As one example, at one point in early 2000, the spot market rates for OC3 private lines from New York City

to Rome were twice as high as those to Milan.) However, this is for links that cost tens of thousands of dollars per month. In the early days of many mass communication services, such as the telegraph or the telephone, when costs and prices were high, we also find prices varying, depending on particular circumstances. (See sections 13 and 14.) Later, decreasing prices were associated first with a move to simple pricing based just on distance. Such rates thus already represent a considerable concession to the desire for simplicity! As a technology evolves, costs drop, and usage increases, simple distance-only rates tended to get replaced by even simpler distance-insensitive prices. This was true for both government monopolies, private monopolies, as well as competitive carriers.

7. The conventional economic argument for flat rate pricing

There are strong incentives for service providers to engage in price discrimination, and increasing opportunities to do so. However, there are also arguments for flat rate pricing. A frequently cited one is that of reduced costs. It is often claimed that in voice telephony, billing costs more than the network. Sometimes the example of the British Post Office reform of 1840 is cited, which introduced a distance-insensitive rates, thereby eliminating the laborious processes of computing correct charges. This reform is then claimed to have eliminated most of the costs of the postal system, making lower rates possible, and stimulating usage. For both the phone system and the mail, there is a grain of truth in these stories, but the full picture is much more complicated, as we will see in later sections. For the Internet, though, it is not clear whether these stories would be relevant, even if they were completely accurate. Transaction costs can often be decreased dramatically online, and there are examples of usage-sensitive charging that appears to operate smoothly for business connections to the Internet (cf. Section 19). Hence the reduced cost of flat rate pricing may not provide a persuasive argument.

A more substantial reason for preferring flat rates is that they reduce transaction costs of the users. In some cases these costs are tangible. For example, the introduction of mailboxes in the U.K. soon after the postal reform of 1840 brought in uniform rate for letters meant that senders could simply drop their stamped letters into those boxes without having to travel to the nearest post office. This offered a clear saving to society that did not appear in postal accounts. On the Internet, such tangible savings are harder to realize. However, there appear to be significant intangible savings, which will be discussed later. First, we discuss some tangible benefits of flat rates to carriers.

For service providers, flat rates do offer substantial advantages, since they represent a form of bundling. Bundling is the strategy of offering several products or services for a single price, and is common. A frequently cited argument for bundling is that it simplifies life for consumers by providing

a single bill and a single point of contact. Such convenience is desirable. If that were the complete explanation, though, bundling might justify extra charges for the added convenience. In practice, though, bundles are almost always priced lower than the sum of prices of their components. For example, the Microsoft Office bundle typically costs about half of what purchasing Word, Excel, Powerpoint, and the other pieces separately would cost. Thus the main justification of bundling has to be something other than convenience.

The main incentive for service providers to bundle is that it allows them to obtain higher revenues. They do this by taking advantage of uneven preferences among consumers for various components of the bundle, and thereby reduce consumer surplus. A simple example illustrates this idea. Suppose that Alice and Bob are on some island that has its only connection to the Internet through an expensive satellite link, and the service provider decides to extract the maximal possible value by charging separately for access to every Web site, but is not able to discriminate between individuals. Suppose that, given the high prices likely to be charged, Alice and Bob are interested just in CNN and eBay, and that Alice is willing to pay \$15 per month for access to CNN, and \$25 for eBay, while Bob is willing to pay \$20 per month for CNN, and \$10 for eBay. If the service provider offers access to CNN and eBay separately, it will maximize its revenues by charging \$15 for access to CNN, and \$25 for eBay, which will produce total revenues of \$55 per month, with Alice paying for both CNN and eBay, and Bob just for eBay. On the other hand, if the service provider bundles access to both CNN and eBay for \$30 per month, then both Alice and Bob should purchase the bundle (even if they have the option of getting access to CNN and eBay separately for \$15 and \$25, respectively), and total revenues will increase to \$60 per month. When the marginal cost of providing access to eBay is less than the \$5 increase in revenues, bundling increases profits for the service provider, and at the same time increases the welfare of Alice and Bob. Note that if Alice's and Bob's willingness to pay were known to the service provider, and it could discriminate, it could charge Alice \$40 per month, and Bob \$30 per month, for total revenues of \$70 per month. Bundling in this case leads to an increase in revenues towards the highest possible (namely the customers' total willingness to pay) without explicitly discriminating between Alice and Bob.

Bundling (which has been studied extensively in economics) is not always more profitable than selling a-la-carte. However, for most reasonable distributions of demands, bundling is better for the service or good provider. (However, mixed bundling, in which the various items are offered for sale separately as well is always better.) In general, the more items there are in the bundle, the greater the gain in bundling (at least for zero marginal costs). That is one reason large carriers that offer multiple

services can obtain advantages in the marketplace through clever pricing strategies.

Flat rate pricing plans are a form of bundling. Again, consider a simple example. Suppose that among the hundreds of millions of Web sites, Alice is interested in just 10, and would like to download 2 MB per month from each. However, she is willing to pay \$0.40 per month for the material from site 1, \$0.80 for that from site 2, and so on up to \$4.00 per month for access to site 10, as shown in Table 7.1. Thus she should in principle be willing to pay a flat rate of \$22.00 per month for access to all the sites. However, on a per-byte basis she is only willing to pay \$0.20 per MB for material from site 1, \$0.40 for material from site 2, and so on up to \$2.00 per MB for material from site 10. Now suppose that Alice’s ISP charges strictly per byte volume. If the price is set at \$1.00 per MB, Alice will pay only for access to sites 5 through 10, and will download 12 MB per month for total revenues to the ISP of \$12.00. If the price is \$1.20 per MB, Alice will pay only for access to sites 6 through 10, download 10 MB per month, and the revenue will be the same \$12.00 per month. Any other price will result in even lower revenues for the ISP. Thus in this example, flat rate pricing produces revenues that are 83% higher than with volume charging.

Table 7.1. Flat rate pricing as bundling

site	willingness to pay
1	\$ 0.40
2	0.80
3	1.20
4	1.60
5	2.00
6	2.40
7	2.80
8	3.20
9	3.60
10	4.00
total	\$22.00

The example above does not produce conclusive proof that flat rate is better for the ISP than volume charging. The gain of 83% shown there holds only if the cost of carrying the additional traffic in the flat rate scenario is negligible, and if the ISP can guess that \$22.00 is Alice’s maximal willingness to pay. If the ISP prices the flat rate service at \$25.00 per month, Alice will not sign up, and revenue will be zero.

Suppose that providing the additional service that flat rate stimulates carries negligible marginal cost. (This is not true for residential modem access to the Internet, since modem maintenance is the

most expensive part of running such a business. However, for cable modem or ADSL services that are always on, this can be almost true.) Let us further assume that, as above, users know what utility they derive from each possible bit they might receive or send. Under those conditions, whether a service provider obtains more revenue from a flat rate or a usage-based tariff depends on the distributions of demands among potential customers. This problem is considered in [FishburnOS], where it is shown that for a majority of what seemed to be a large class of realistic demand distributions, flat rate produces higher revenue for the service provider.

The point of the discussion above is that even in the conventional model with simple consumer utilities, it is often more profitable for the service providers to offer flat rate service. As it turns out, there are other factors that make flat rate pricing even more compelling.

8. The strong public preference for flat rate pricing

While charging according to usage has advantages, it is disliked, and people are willing to pay extra to avoid it. The strength of the preference for flat rate pricing is illustrated by the history of America Online:

What was the biggest complaint of AOL users? Not the widely mocked and irritating blue bar that appeared when members downloaded information. Not the frequent unsolicited junk e-mail. Not dropped connections. Their overwhelming gripe: the ticking clock. Users didn't want to pay by the hour anymore.

...

Case had heard from one AOL member who insisted that she was being cheated by AOL's hourly rate pricing. When he checked her average monthly usage, he found that she would be paying AOL more under the flat-rate price of \$19.95. When Case informed the user of that fact, her reaction was immediate. 'I don't care,' she told an incredulous Case. 'I am being cheated by you.'

[Swisher], pp. 160-162

When faced with a choice of pricing plans, people often pay more for a flat rate plan than if they signed up for usage-based plans. This is true for a large fraction of residential Internet users, and not just the customer of AOL that Steve Case talked to. Many ISPs do offer hourly pricing plans, and those typically would be financially advantageous for many of the subscribers who are on the unlimited plans.

The willingness to pay extra in order to avoid metered services has been observed elsewhere. Perhaps the first well documented large scale experiments in this area were those carried out by the Bell System in the 1970s. By that time the United States had had unlimited residential local phone calling for a flat monthly rate for over half a century. (There were exceptions to this rule, for example New York City and Chicago, but most of the country was on flat rate plans. The fascinating story of how this came about, even while most of the world moved to metered service, is summarized in Section 14.) In the Bell System experiments of the 1970s, customers were usually offered a choice of three plans. The least expensive (in terms of the basic monthly rate) allowed for only a few local calls without additional charges, and then had high fees for additional calls. The most expensive plan provided unlimited local calling. It was a great surprise when most of the people decided to stay with flat rate plans. This even included the majority (typically over 60%) of the people who were making practically no phone calls, and would have benefited the most from the usage sensitive tariffs. The results of these experiments are described in [CosgroveL, GarfinkelL1, GarfinkelL2], and are also discussed in [FishburnOS].

The Bell System experiments elicited extremely strong negative reactions from customers. Some states enacted laws forbidding carriers from discontinuing flat rate plans. This reaction caught both AT&T and FCC officials by surprise. Both parties were convinced of the irrationality of flat rate plans and the advantages of metered service. Yet, had they been aware of AT&T's experience 70 years earlier (briefly described in Section xxx), they might have been prepared for this reaction. Investigations of the experiments of the 1970s uncovered three main reasons for the flat rate preference:

(i) Insurance: It provides protection against sudden large bills. (What happens if my son comes back from college, and starts talking to his girl friend around the clock?)

(ii) Overestimate of usage: Customers typically overestimate how much they use a service, with the ratio of their estimate to actual usage following a log-normal distribution. (This is a general phenomenon. A recent study found that email users reported receiving about twice as many messages as they actually did [Kanfer].)

(iii) Hassle factor: In a per-use situation, consumers keep worrying whether each call is worth the money it costs, and it has been observed that their usage goes down. A flat-rate plan allows them not to worry as to whether that call to their in-laws is really worth \$0.05 per minute.

Of these three reasons, the first appears to be the most important. People are risk averse. The famous study of inner city residents in Camden, New Jersey, by Mueller and Schement [MuellerS] found this to be a major reason that many households had cable TV, but no phone service.

Since the 1970s, much additional evidence has accumulated of the strong preference for flat rate plans. People are routinely willing to pay more to avoid metered billing. For example, a large fraction of residential Internet users would save by paying their ISP's hourly rates instead of the standard flat rate that they choose. Even when true flat rate is not available, block purchase plans (in which users pay a fixed fee for a large block of calling time) have been popular. Consumer usually buy considerably larger blocks of time than they use [Palmeri, MitchellV].

Recently, further detailed evidence of the attractions of flat rate plans has been obtained in the carefully controlled INDEX experiment. This experiment, described in [AltmannRV, EdellV], tested users' willingness to pay for various Internet access options. Most of the INDEX investigators appear to conclude from the experimental data that differentiated services and usage-sensitive pricing would be better for both ISPs and users (cf. [EdellV]). However, one can also interpret the data as supporting flat rates. One argument in this direction is based on the dramatic effect that metered billing has in decreasing usage, and is presented in the next section. Right now we mention another argument. In some phases of the experiment, participants were offered a choice of paying a flat fee to free themselves from the constraints of metered billing. It was discovered that on average, these participants were willing to pay considerably more in flat fees than through metered billing [Chu]. Their willingness to do so in a commercial setting would probably be much greater than it was in the INDEX experiment. In INDEX, they had excellent and immediate feedback on their usage (and thus on their real or potential bill), easy control over their options and could change their pricing plan each week. In a commercial setting, where they might not see their charges until their credit card bill arrived six weeks later, they would likely be willing to pay considerably more to avoid unpleasant surprises.

Of the reasons for the flat rate preference that were uncovered in the Bell System trials in the 1970s, the third reason is part of what Nick Szabo has called "mental transaction costs" [Szabo]. Computer processing power and bandwidth are growing and getting less expensive. However, our personal processing power is not increasing. Human attention is an ultimate scarce resource. (See [Goldhaber] for an extreme view in this direction.) Psychologists have shown that our short term memories can hold only three to five chunks in short-term memory tasks. Do we want to reserve any of them for pricing considerations?

9. The dramatic effect of usage sensitive prices

Usage sensitive prices are effective. This is known very well, and there are established methodologies and a huge literature on estimating elasticities of demand for telephone service (see, for example,

[CracknellK]). There are graphical demonstrations of the effects of time-of-day variation in usage sensitive prices on telephone use in [deSolaP1, Mitchell1]. For residential Internet customers in Europe, the graphs in [Blair] show spikes in usage when less expensive evening rates for local calls begin to apply. Even mail usage is sensitive to prices, as is demonstrated by the data in [Wolak].

The main problem of usage sensitive prices on the Internet is that they are often too effective and their effects are sometimes undesirable. In particular, there appears to be a very sharp threshold effect, in which any non-zero price reduces usage drastically. For example, the protest against AOL usage sensitive charges described in the quote from [Swisher] in Section 8 led AOL to introduce flat rate pricing in October 1996. This led to a tripling of usage per subscriber, as is shown in Table 9.1. This tripling took about a year, but chances are it took so long only because the AOL network had to be expanded to accommodate the new demands. (In early 1997, AOL became known as “America Off-Line,” and had to face lawsuits by disgruntled subscribers not able to log in. It had to stop recruitment of new users, and concentrate on infrastructure improvements.) By comparison, Europeans, who pay for local calls, have not increased their usage anywhere near as much. As Table 9.2 shows, their usage is about one third the U.S. level.

Table 9.1. Time spent online by AOL customers

calendar quarter	average time per customer per day, in minutes
2Q1996	13
3Q1996	14
4Q1996	21
1Q1997	31
2Q1997	38
3Q1997	40
4Q1997	41
1Q1998	46
2Q1998	44
3Q1998	47
4Q1998	48
1Q1999	55
2Q1999	52
3Q1999	55
4Q1999	57
1Q2000	64

In the AOL example, elimination of usage sensitive pricing in favor of flat rates resulted in a huge increase in usage. Projections based on marketing studies in the U.K. suggest that flat rates there would

have a comparable effect, tripling the amount of time British subscribers spend online [Durlacher]. That suggests the lower intensity of Internet usage in Europe is caused by pricing, not by culture. Similarly, the differences in the amount of time that people spend on the phone in different countries visible in Table 1.1 are not just a reflection of culture, and in particular of any special propensity of Americans to waste time on the phone. It is instead a reflection of the interaction of culture and pricing. Surprisingly, it appears that no careful study has been done of the effects of switching from flat rate local call pricing to metered rates in various countries. (See [Kraepelien] for the best listing of data on this subject that I have been able to find. Unfortunately it is extremely limited and lacking in detail.) It appears that imposition of metered rates has always led to a substantial drop in usage. For example, when British Telecom (then part of the British Post Office) started charging for local calls in 1920, there was a sudden drop (starting even before charges came into effect) by about a third in the number of calls per line, followed by a long period of relatively steady decline [Cracknell1].

Table 9.2. Time spent online by French residential Internet (not Minitel) users

month	number of users (millions)	average time per customer per day, in minutes
Sep. 1996	0.15	8
Sep. 1997	0.40	15
Jan. 1998	0.54	14
Apr. 1998	0.70	15
July 1998	0.80	15
Oct. 1998	0.96	16
Jan. 1999	1.28	17
Apr. 1999	1.50	17
July 1999	1.64	17
Oct. 1999	1.93	17
Jan. 2000	3.03	16

Many more examples of how metered rates suppress usage, and flat rates increase usage, will be presented later, especially in Section 11. Decreasing usage might seem desirable, to eliminate waste, decrease congestion, etc. For example, the U.K. switch to metered local calls in 1919 was prompted by the large financial losses that the phone system was sustaining, losses that were being covered by profits from the mail. The result of the change, as seen in the statistics in [Mitchell2], was that the number of telephone lines continued to grow, even as the total number of lines decreased. The British government may have felt that this was a worthwhile tradeoff, allowing more people to have basic service.

Decreasing or at least limiting phone usage might have seemed desirable in the phone systems of

the U.K. and other countries in 1919. In retrospect, Table 1.1 raises a question about the wisdom of that move for the long run, but we do not have a careful study evaluating its effect. There were certainly good arguments for the switch to metered rates, much stronger than today, as is described in sections 4 and 12. However, on the Internet, we are in a different environment. In a rapidly growing market, service providers strive to attract users. It is noteworthy that AOL resisted the switch to flat rate pricing in 1996, fearing (correctly) that it would increase usage. However, just as Dr. Strangelove and The Bomb, AOL learned to stop worrying and to love flat rate. The Feb. 2, 2000 press release quoted in Section 3 also states that

For the first time, AOL subscribers now are averaging more than one hour online daily – 63 minutes this month – underscoring that the AOL experience is becoming increasingly valuable and necessary to its members’ everyday lives.

AOL’s current business model is to become yet more “valuable and necessary to its members’ everyday lives.” Its plan for the next 5 years calls for tripling the amount of time that individual subscribers spend online per day [Hansell].

The advantages of flat rates in encouraging Internet use have been recognized widely around the world. Section 17 describes how countries such as Germany, Japan, and the U.K. are reintroducing flat rates (often in a limited form, just for Internet access) precisely to promote Internet usage.

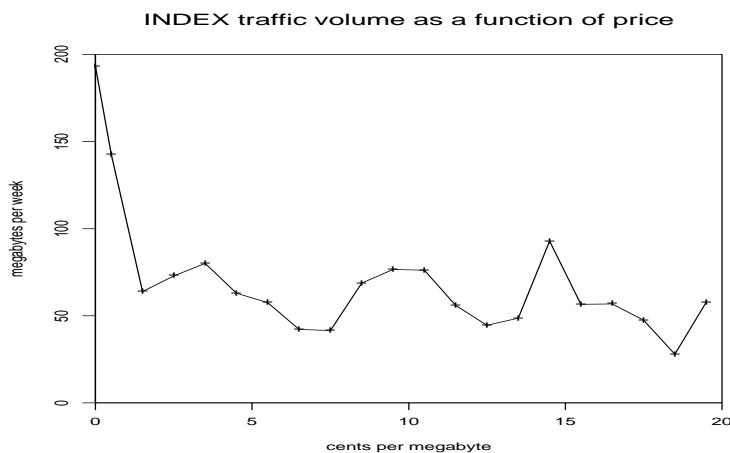


Figure 9.1. Data received by participants in the per-byte charging part of the INDEX experiment as a function of price.

The decreases in usage discussed above that are caused by metered rates can in principle be ex-

plained in standard economic terms. All that happens is that the transactions with low value to the user are not carried out, which increases aggregate welfare if that value is less than the cost to the network. The problem with this conventional explanation is that many of the observed decreases are considerably larger than one would normally expect. That this is indeed a problem, and that a psychological explanation is more appropriate, can be seen in the data from the INDEX experiment, mentioned in the previous section, and described in more detail in [AltmannRV, Chu, EdellV]. During one phase of that experiment, participants were subject to straight per-byte charges. In the initial week, all 56 participants in that phase had free service, and downloaded an average of 193 MB each. During each of the following six weeks, they had to pay for all traffic to their homes, with prices chosen according to the principles of statistical experiment design, and taken from the range of \$0.001 to \$0.20 per MB. (It is worth noting that residential modem customers typically download about 60 MB per month. Thus, if they are charged \$20 by their ISP for unlimited service, they are effectively paying about \$0.33 per MB. Cable modem or DSL customers typically download 5 to 10 times as much, but usually pay between \$40 and \$60 per month, for an effective price of between \$0.07 and \$0.20 per MB. Thus the prices in the INDEX experiment were not very high.) As is shown in Fig. 7 in [EdellV], even low prices led to a big drop in usage compared to the week with no charges. However, that bar chart aggregates data to such an extent that it is hard to tell what the effect of very small charges is. The INDEX experimenters have kindly supplied me with the raw data for that chart, and I have processed it to produce Fig. 9.1. (The interpretation and presentation of the data are strictly mine.) The entry for the price of 0.5 cents per MB is the average of the subject-weeks where the subject faced a price of between 0.1 and 1 cents per MB, the one for a price of 1.5 cents per MB is the average for prices between 1.1 and 2 cents per MB, and so on. Each data point corresponds to an average of 14 subject-weeks, so that experimental error was substantial, and accounts for the jagged line. (There was clearly nothing special about prices between 14 and 15 cents per MB to induce a large jump in usage!) The main point is that even prices below 1 cent per MB led to a giant drop in usage (to an average of 143 MB per week from the 193 MB during the week of free service). Prices between 1 and 2 cents led to a further drop to an average of 64 MB per week. Further increases had a much smaller effect on usage. Yet even at 2 cents per MB, the usage of 200 MB/week that was observed during the free week would have cost just \$4 per week. It seems more reasonable to attribute this behavior to a psychological aversion to metered rates than to a utility function that has an unusual shape.

10. Dynamic effects of flat rates

Whether to offer flat or metered rates in a competitive environment is a business decision. One way to select the preferred course is through a traditional accounting analysis. Find out how much potential customers are willing to pay extra for the flat rates. (Either run an experiment like INDEX, or do market trials, focus groups, and other standard methods of marketing.) If this amount is large enough to compensate for the inefficiencies introduced by a move away from metered pricing, give the customers what they like, namely flat rates. If the willingness to pay for flat rates is too low, offer only measured rates. While such an analysis seems appealingly rational, it is likely to be seriously misleading in an environment such as that of the Internet, where rapid growth is the rule. The problem is that this analysis is static, and ignores both network effects and temporal evolution. To properly evaluate the choice between flat and measured rates, one has to take into account the indirect effects of those rates on other users, as well as on future usage of a particular customer.

As Section 4 points out, the value of a network is based not just on the number of users in it, but, even more, on the intensity with which it is used. If you are connected to a network such as the Internet, you can be reached by others, and you may contribute content to the network. You are also more likely to try new services and products if you are online often. Thus your presence makes the network more valuable to others. (This argument applies much less to a broadcast network, but as I have argued in Section 2, the main value of the Internet is in the interactivity it provides.) Hence some of the costs of your usage can justifiably be spread over other participants. Thus the loss in efficiency that flat rates cause is smaller than it might seem at first, and the benefits from increased usage that flat rates stimulate are greater.

A related reason that the static analysis of willingness to pay for flat rates is likely to be misleading is that it neglects the dynamic aspects of the Internet connectivity market. Network service providers have every incentive to persuade customers to move up to higher quality connections. The first step is to move them from dial modems to cable modems and ADSL links. The next step will be to persuade them to pay for fiber connections, on which bandwidth can be increased step by step. Given the progress in technology, providers are forced to plan on persuading their customers to trade up. A fixed link is too likely to be commoditized, and unable to offer the services that people come to expect. Who is most likely to want to move to a higher bandwidth connection? Is it an American modem customer who pays flat rates for local connectivity and Internet access and is online about an hour per day (Table 9.1)? Or is it a French modem customer who pays by the minute, and is online for 17 minutes per day (Table

9.2)? Or is it a French Minitel customer of a few years ago, who paid by the minute and was online for under 3 minutes per day (Table 16.1)?

The above two effects cannot be quantified yet, with our current state of knowledge, but they appear substantial. They certainly did play a role in the evolution of pricing of various communication services, as we will see later, and they argue strongly in favor of flat rates.

In early 2000, the company Broadband Digital Group (<http://www.bdgcorp.net/>) announced a plan to offer free or low cost DSL service in the U.S. Their pricing strategy is explicitly based on the expectation of trading their customers up to better and more expensive services. 144 Kbps service will be offered for free (aside from modem and installation charges) for those customers who agree to accept ads, or for \$9.95 per month without ads. There is no hope that these charges will cover the costs. However, “[s]peeds of up to 384 Kbps will cost \$19.95 a month without ads, and speeds of up to 1.54 Mbps, \$34.95 a month without ads,” according to news reports. The hope is that enough of the customers who sign up for the free or low cost versions will trade up to the higher quality to make the venture successful. Whether this particular attempt will succeed, or will be seen as just another instance of the “irrational exuberance” of the financial markets remains to be seen. However, the basic approach of migrating customers to better services appears the right one. It is worth noting that this approach is not novel. Even a few decades ago in the U.S., party lines (in which several households shared the same line) were common. They were usually thought of, by both phone company officials as well as the public, as a way a common way to reduce costs. However, others advocated a different view of this service:

There seems to be two distinct fields for two-part lines: 1st) To provide a service which shall be so limited in amount and of such inferior character, that—while (at a low price) it will attract new subscribers—will not satisfy them permanently, but will act an an educator and feeder for the more profitable grades of service. To accomplish its purpose, and extremely low rate must be affected, entailing to this company considerable loss, which loss may be considered a charge against procuring of new business. The object of this service will not be accomplished unless the service is unsatisfactory, when once taken. It therefore requires that enough subscribers be placed upon a line to make them dissatisfied and desirous of a less limited and better service.

memo from AT&T Chief Engineer

Joseph P. Davis, 6 February 1899

as quoted on p. 48 of [Fischer2]

This represents a combination of a dynamic view of a communications service and the “damaged goods” approach discussed in Section 5.

11. Miscellaneous examples of pricing trends

In pricing, there is a continuing conflict between two basic tendencies. The drive for economic optimality argues for sophisticated pricing strategies in all situations. On the other hand, consumer preferences are for simplicity at all times. Technology is the crucial third ingredient, offering a range of choices. This section considers pricing trends in a variety of goods and services, to illustrate how the evolving technology is affecting where the balance between the drive to optimize and the reluctance to be optimized is drawn.

The business-to-business market is different from the business-to-consumer market, and the focus of this paper is on the latter, where individual preferences matter the most. There is no sharp dividing line between the two markets, but there is a substantial distinction. For example, McDonald’s offers free sugar to its customers, and builds the cost of this service into the price of the coffee. It would surely be more efficient in terms of allocating resources to charge for each packet. It would also be fairer, in that customers who do not use sugar are subsidizing those who do. (It would also be healthier for the customers!) Yet that is not done, and the customer desire for simplicity as well as McDonald’s desire to keep transaction costs low lead to “wasteful” practices. On the other hand, McDonald’s buys its sugar in bulk, and undoubtedly has purchasing experts who play off various suppliers against each other, arrange long-term contracts, and possibly even trade on the sugar futures markets. There is a spectrum of business decisions on how far to optimize. The general tendency for businesses appears to be to optimize when the optimization can be done by dedicated specialists (such as the sugar buyers at McDonald’s). When the optimization requires many small actions by large groups of employees, the tendency appears to opt for simplicity, as we will see.

Simple pricing, involving strategies such as bundling and flat rates, is very common. Table 2.1 shows that we spend far more on simple purchases of recorded music than we do on videotape rentals or on admission to movie theaters. Each of those categories of spending is dwarfed by the purchases of consumer entertainment equipment.

If flat rates are so great, why don’t we have more of them? Flat rate airplane travel is clearly impractical. Most people do not make any air trips in a given year, and prices vary too wildly. How

about something like Coke? It is inexpensive and bought frequently, so why not have flat rate Coke plans? Well, we sort of do. At many companies in the U.S., especially in Silicon Valley, free soft drinks are provided to all employees. Also, on cruise ships, food and soft drinks are included in the basic fee. However, we do not see Coca Cola offering unlimited access to vending machines for a flat fee. There is nothing to stop it (or, more precisely, the local distributors or vending machine owners) from making such arrangements. Conceivably some may do so in the future, especially as vending machines become more intelligent and can communicate with smart cards or other devices that can authenticate users. Still, it is easy to think of many reasons why vendors might be reluctant. One relates to what economists would call arbitrage or incentive compatibility (terms that are avoided in this paper together with all other professional economics language). How would a vendor stop a flat rate customer from taking enough cans of soda for a whole party? It is easy to imagine a single user emptying a whole vending machine. Now this might not be a serious problem (newspaper vending machines do not seem to suffer from it too much), and one can also put in controls, such as limits on the number of cans a user could take in a day. So perhaps, as vending machines become smarter and get networked, there will be flat rate Coke plans. (This is an instance where technology might make it easier to implement flat rate plans, and not just enable quality and price differentiation strategies.) However, we do not see them yet.

In more expensive and less frequently purchased goods, we see few signs of simplicity. Saturn's no-haggle policy elicits positive comments from customers, but has not overturned the prevailing sales policies of car dealers in the U.S. (In other countries, there is less flexibility since car makers tend to have strict controls over their dealers, so bargaining is less common.) Even in department stores, periodic sales (which are a form of price discrimination, as was shown by Varian [Varian1]) prevail. In the early 1990s, when Sears underwent a major restructuring, it attempted to give up sales in favor of "everyday low pricing." However, this policy apparently could not be sustained, and Sears has gone back to the old style of selling.

Airlines not only have several classes of service, but often have separate check-in lanes for each. On the other hand, supermarkets and banks do not have separate checkout lanes for people who might be willing to pay 10% extra for faster service. Even airlines tend to have a single class of service on short flights.

We next consider in more detail pricing trends in several industries.

(a) Airline yield management.

Airline pricing is something that almost everybody loves to hate. Constantly changing prices, inconvenient restrictions, passengers in adjacent seats paying fares that differ by a factor of 10, and poor quality meals (if any) in coach class are the main annoyances. Yet most economists who have studied the system applaud it for its efficiency in satisfying the demand for fast transportation. Average fares (adjusted for inflation) are going down, although slowly, since technology is rather mature. At the same time, the industry is nicely profitable. For passengers, while the average fares are declining, seats are usually available even at the last moment (although at high prices). The airlines are using a sophisticated yield management system that adjusts prices to meet demand. Various industry analysts report that “business travelers take 1/3 of seats, but provide 2/3 of revenue,” or that “the 10 percent of passengers who pay full fares account for 40 percent of revenues.” The system is designed to charge according to willingness to pay, and appears to be doing this effectively.

Airline pricing, even in the U.S., used to be simple. Fares were regulated by the government, price competition was unheard of, and airline rivalry was confined to amenities and flight convenience. Deregulation and technology have changed this. Competition has added an extra spur to the incentives for price and quality differentiation, while technology has provided the tools to implement such schemes. (It is worth emphasizing the role of technology, especially that of combinatorial optimization algorithms, which received a boost from the invention of the Karmarkar interior point method and its application to airline scheduling. Today, even the supposedly fixed costs are becoming less fixed as a result of flexible capacity assignment policies.

The efficiency of airline yield management systems is enhanced by non-transferability of tickets. Were these tickets transferable (as bus or rail tickets are), there would be opportunities for arbitrage, and 10-fold differences in price for essentially the same good could not be sustained. That is why airlines are so happy to enforce document checks. Thus the legal nature of the contract (a personal service as opposed to a transferable good) is leading to greater price discrimination. The main effect of strengthening of intellectual property protection laws will be again in this direction, towards greater price discrimination.

Although yield management systems are constantly improving, the attractions of simple fares have been noted. Even in the competitive airline market in the U.S., there have been several attempts to introduce simple fare structures. The most notable one was that of People Express. There were also at least two attempts by Americal Airlines to introduce simple schemes. All

failed, and this suggests that in expensive and infrequent purchases like those of airline tickets, the demands for maximizing revenues is sufficient to overcome the public's dislike of the methods that are used.

Other industries are following in the footsteps of the airlines. Hotels, cruise ship lines, and even golf courses are introducing such systems. It clearly does serve a purpose. However, customer reactions are seldom positive. For example, a recent news story about British railroads spent as much space complaining about the unpredictable fares as about the deterioration in service levels [[Cowell].

(b) Lexis-Nexis database.

The standard pricing plan for access to the Lexis-Nexis databases of legal and news information used to be metered, with charges for connect time, searches, and so on. This plan was roundly disliked by customers. Starting a decade ago, a modified flat rate plan was introduced. Client companies are given unlimited access to some agreed databases for a year for a fixed fee that is negotiated, and depends on the size of the firm and on prior usage. At the end of the year, the price (and databases that are covered) are renegotiated. This plan is very popular. Although precise statistics have not been released, it apparently is used by the vast majority of the larger customers.

When a Lexis-Nexis customer switches from metered billing to the modified flat rate plan, usage on average goes up by a factor of 2.5 (public presentation by Keith Hawk of Lexis-Nexis at the SCOUG symposium on April 30, 1999).

One can speculate that one reason Lexis-Nexis has been willing to move to the modified flat rate plan is that their marginal costs of providing connectivity and search capability to their customers have declined. In the beginning, their costs may have been dominated by those of maintaining the complicated mainframe center that could accommodate their database. However, improvements in technology have undoubtedly reduced costs of computers to the point that total Lexis-Nexis costs are dominated by the first-copy costs of updating the databases.

There have been arguments that pay-per-view schemes in pricing of electronic scholarly journals would offer advantages to publishers and readers [ChuangS]. However, the overwhelming trend is towards subscription fees, bundling, and consortium licensing, as was predicted in [FishburnOS, Odlyzko1], for example. There is now extensive experience, including results from the carefully conducted PEAK experiment [GazzaleMKM], which tested a variety of pricing plans.

One of the principal findings was that metering access to electronic information drastically decreases usage. Elsevier, the largest scientific and technical publisher, and the main provider of material for PEAK, has modified its pricing policies in light of the evidence from that trial, so as to emphasize subscriptions, supplemented by sales of individual articles where necessary:

[Elsevier's] goal is to give people access to as much information as possible on a flat fee, unlimited use basis. [Elsevier's] experience has been that as soon as the usage is metered on a per-article basis, there is an inhibition on use or a concern about exceeding some budget allocation.

[Hunter]

(c) Water.

Most utilities, such as water, electricity, and gas, are metered. There are many arguments for such metering. Technology is usually not advancing rapidly, and often there are increasing costs (such as those of dealing with pollution). Typically there are also obvious and substantial marginal costs associated with the provision of the service. However, it is interesting to observe that in some cases flat rates persist, and users are strongly attached to them.

When municipal water supply systems were widely developed during the 19th century, initial pricing tended to be flat rate, based on various measures of expected usage (such as the number of rooms in a house, [Jordan]). Eventually most of these systems converted to metering, but against strong opposition from users.

The conversion to metered pricing for water in the U.S. is not complete. Surely the most prominent water supply system with flat rates is that in New York City. Commercial establishments were converted to usage-sensitive charging decades ago. Residential customers are only now converting, and only because of intense pressure from the city. The basic problem is apparently in both supply and disposal. New York City is in the process of completing its fabled Third Water Tunnel, which will provide much needed spare capacity and also redundancy in case one of the other, century-old tunnels fails. There is also a shortage of waste-treatment capacity, and construction of new plants for that purpose would not be popular in any neighborhood. Hence, in order to cope with rising demand, the city government is resorting to pricing to limit usage. This move is extremely unpopular. Several deadlines have slipped, and the latest move is a doubling of the flat rates. (More details and references to come.)

What is the effect of metering water in The Big Apple? Usage drops by between 20 and 50%. (The decrease is larger for single-family housing, where the cost is born directly by the occupants, and smaller in apartment houses, where the allocation of costs is more involved.) Usage remains high, so it is unlikely that people refrain from showering. However, they presumably fix leaks faster than they did with flat rates, or do not water their gardens as much.

Why do New York City residents resist metered water? The main complaint is that costs would go up. In practice, costs go down. (Overall, since the city is recovering the costs of its service, one would expect that total bills would stay about constant, as the predominantly fixed charges get redistributed over smaller usage. In the long run, not building new facilities should serve to keep charges from rising.) However, as with so many other complaints about metered rates, what the residents may really have in mind is not just the purely financial costs they bear, but the mental costs involved in metering.

The history of water supply pricing shows the strong attachment that people place on flat rates. It also shows another feature that is of likely relevance for the Internet, namely of having a single high quality for all supplies. All municipal water is treated to a standard adequate for drinking. This is a monumental waste, since most of the water is used for flushing toilets, or watering gardens, where much lower standards would suffice. However, the extra costs of having two separate systems, and of having to keep track of them, have prevented (with some exceptions) development of water systems with two quality levels. It does not seem unreasonable to suggest that a similar logic will prevail on the Internet, and that the costs of having differentiated services will be seen as higher than the inefficiency of having a single high quality of transmission for all traffic.

(d) Electricity.

The electric power industry even in its early days paid careful attention to issues of pricing and usage patterns. A large fixed investment was required to start service. Furthermore, the initial demand was for lighting, and so lasted only a few hours each day. This led to extensive discussions and studies on how to price electricity. An excellent short summary is presented in [Friedlander3], and more detailed coverage in [HausmanN1, HausmanN2, Hughes, Platt].

More discussion will be provided in the next version of this work. Right now we note that the first electric power systems did have flat rate service (charged by the number of light bulbs). However, metered services took over quickly. Also, we note that in general, industrial users pay

lower rates than residential ones.

(e) Congestion tolls, garbage collection, and other examples.

Road congestion is a perennial problem. It is an attractive candidate for applying pricing measures, since it is well established that building more highways only stimulates more travel, which brings congestion back. (For further discussion and references, see Section 22.) In the past, it was difficult to use pricing for car travel. Advances in technology, though, are making this increasingly feasible, and many experiments are being tried or are in the planning stages. It appears too early to tell just how effective this will be.

Some results have been ambiguous. For example, on a section of Highway 91 between the center of Los Angeles and Riverside, drivers have a choice of the free public lanes or toll lanes constructed by a private company. This is an instance of Paris Metro Pricing [Odlyzko5]; there are no guarantees of performance, and drivers have to commit themselves to one of the alternatives without knowing what the road conditions are. Prices vary dynamically. Reports on the effects of this scheme are mixed. It appears that the private contractor is not making money, and congestion is seldom encountered on either the free or the toll lanes. Instead, it has reportedly moved to the area where traffic separates into the two streams. A similar scheme exists in Houston. There are three parallel roads from the downtown area that go past the airport, one of them a toll road. In that case it is reported that the toll road does offer consistently lower congestion, leading to less variation in travel times and pleasanter drive.

Another area where usage-sensitive pricing has long seemed appealing is in garbage collection. The usual method for charging in the U.S. has been either to cover it out of real estate tax (at least for residences, as commercial establishments often have to deal with carters on their own), or else through flat rates. When shortages of landfills appeared in the 1970s and 1980s, there was an attempt at wider use of usage-sensitive pricing. They were effective, but, as so often happens, not all the effects were positive. Some of the reductions were in the physical volume of the garbage, not its weight (the result of the “Seattle stomp”). There was also an increase in illegal dumping and littering [FullertonK].

12. Mail

Postal systems are ancient, and there is a voluminous literature on their history, technology, politics, and economics. This literature all by itself illustrates the conflicting drives for price discrimination,

efficiency, and simplicity. that have shaped the evolution of communications.

Table 12.1. British Post Office volume and financial statistics

year	letters (millions)	revenue (millions)	profit (millions)
1711		\$0.56	\$0.45
1760		1.15	0.42
1770		1.53	0.78
1780		2.09	0.68
1790		2.86	1.66
1800		5.42	3.60
1810		9.28	5.69
1820		10.96	6.93
1830		11.33	6.52
1839	75.9	11.95	8.17
1840	168.8	6.80	2.50
1841	195.5	7.48	2.81
1842	208.4	7.89	3.00
1843	220.5	8.18	3.20
1844	242.1	8.53	3.60
1845	271.4	9.48	3.81
1846	299.6	9.82	4.13
1847	322.1	10.91	4.92
1848	328.8	10.72	3.70
1849	337.4	10.83	4.20
1850	347.1	11.32	4.02
1851	360.6	12.11	5.59
1852	379.5	12.17	5.45
1853	410.8	12.87	5.87
1854	443.6	13.51	5.98
1855	456.2	13.58	5.33
1856	478.4	14.34	6.04

People have been exchanging messages over long distances for thousands of years even without the benefit of writing or a postal service. The desire to communicate was strong enough that messages would be passed along chains of messengers, however haphazard that process might be. Even in the 17th century, English correspondents writing to colonists in North America would address their letters as in the example quoted on p. 226 of [Cressy]: "... at Malden in New England, to be left ... at Mrs. Mary Ardell's at her house in the Town Dock, over against Mr. Thomas Clark, brazier, in Boston." Widespread illiteracy did not mean that written correspondence was useless, since better educated neighbors were often employed to write or read messages. (Similarly, people today can benefit from

the Internet without having Internet access themselves.) Still, organized commercial postal systems were slow to develop.

The earliest postal systems go back to ancient antiquity in places such as Persia and China. As far as is known, they were primarily or even exclusively for government use. Even in more recent times, the first postal systems to be widely used for commercial purposes were originally primarily for government use. (Such was the origin of the famous and extensive Thurn und Taxis system in Western Europe. It started during the Renaissance and in some places endured into the second half of the 19th century.) With the growth of commerce, and the need for economy in operation, these systems were soon extended to carry private letters. Some were operated directly by governments. Others were government-enforced private monopolies, such as that of the Thurn und Taxis family. These monopolies allowed governments to monitor mail communication, and, as traffic grew, also (and eventually of much greater importance) to obtain large profits. The mail business was usually profitable enough to enable the contractors to get rich as well. The contributions of the Thurn und Taxis family to the arts were based on money collected (or extracted, as many have called it) from their postal patrons.

The extraction of profits from postal services may well have reached its highest recorded level in Britain in the early 19th century. That country was desperately searching for money to pay for the Napoleonic and earlier wars and the debts from those wars. The mail turned out to be a bountiful source of funds. Large profits had been made throughout the 18th century, and they increased markedly in the early 19th as rates were raised and economic development of the country led to increased demand for mail service. In 1839, total revenues of the monopoly government postal service were \$11.95 million. The costs of providing the service were \$3.78 million, for an enviable profit margin of 68%. (See Table 12.1.) Proper accounting would have shown even higher profits, since expenses included items having little or nothing to do with postal services. For example, the Duke of Marlborough was paid \$25,000 annually by the Post Office as a reward for his ancestor's military achievements over a century earlier (p. 80 of [Robinson]).

The Post Office profits amounted to over 3% of the British government budget. (Almost 60% of the government budget was for payment of interest on the national debt, so maintaining creditworthiness had paramount importance. Income from mail service was a very substantial fraction, about 8%, or the discretionary budget.) In spite of the high postal charges, the British were active mail users, and were spending about 0.5% of GDP on postage. This was about twice the fraction of GDP that Americans were devoting to this purpose (see Table 3.1). The high profits from the British postal monopoly were accompanied by a complicated pricing scheme, shown in Table 12.2. The "single letter" of Table 12.2

(as well as of Table 12.3, which shows a few snapshots in the evolution of pricing of U. S. postal services) refers to a single sheet of paper, the standard unit for pricing mail services then. One of the key tasks for postal personnel was to inspect these “single letters” against a light, to ensure there were no enclosures, as those incurred additional charges. The average cost of a letter in the U. K. in the 1830s was about \$0.16, roughly two hours of labor for an able-bodied worker. “Single letters” to India often cost most than a pound, or two weeks’ wages (and sometimes took six months to reach their destination).

The extortionate prices charged for private letters in the U.K. in the 1830s were largely a tax, and were recognized as such. Of the small part that went for operations, much was diverted into subsidies for some privileged forms of mail. Franks (officially letters sent by Members of Parliament and other government officials on business, but much abused for private purposes) were estimated at about xxx million pieces a year. Newspapers, of which about 30 million copies were thought to be delivered annually in the late 1830s, traveled free. The official justification was that since these newspapers had already paid their own heavy taxes, they should not have to pay another fee in the form of postal charges. Ulterior motives apparently included issues of political influence and control of the press. In any event, even the internal operations of the British Post Office were full of cross-subsidies. There was practically no relation between costs and prices.

High prices went along with low volume. Readers of Jane Austen or other early 19th century novelists may have the idea of a smoothly functioning and universally available postal system. That was true to some extent for the rich, and especially those rich who lived in large cities, as parts of London had 10 or even 12 daily deliveries. On the other hand, large sections of the country had little mail service. The volume of correspondence was low. Fewer than five letters were delivered per person per year.

Table 12.1 presents some the key statistics of the U.K. postal system around the time of the famous reform of 1840. Accurate statistics for volume prior to 1839 are not available, but the roughly constant revenues and profits suggest there was little growth. The reform of 1840 inaugurated a period of rapid growth. This reform introduced the Penny Post, in which any letter up to half an ounce in weight could be sent anywhere in the country for a (British) penny, approximately \$0.02. The 1840 reform had a profound influence throughout the world, and affected many services aside from the mail. The person usually credited with single-handedly fomenting this revolution is Rowland Hill. Hill came from a family active in reforms during one of the great reform periods in Western history. He did not have any experience with the postal business, and no cooperation from the officials running it. Still, he

Table 12.2. British Post Office rates for domestic “single letter” in 1837

distance	price
up to 15 miles	\$0.083
16-20 miles	0.104
21-30 miles	0.125
31-50 miles	0.146
51-80 miles	0.167
81-120 miles	0.188
121-170 miles	0.208
171-230 miles	0.229
231-300 miles	0.250
301-400 miles	0.271
401-500 miles	0.292

managed to prepare an influential study that he published in 1837 under the title “Post Office Reform; Its Importance and Practicability.” (It is also available in [Hill].) Hill’s pamphlet and his tireless agitation (aided by accidental political factors) led to the 1840 reform. Hill oversaw the implementation of his reform in the face of general resistance of postal officials. As is customary today, his legacy has attracted the attention of revisionist historians (cf. [Daunton]). Some of his proposals may not have been as original as he or later followers claimed, and he was a poor administrator. (The last was claimed by Anthony Trollope early on, based on Trollope’s first-hand experience working for the Post Office). Further, Hill’s postal reform could only be carried out as part of a more general reform movement that led to more democracy in Britain, lower tariffs, and lower taxes. Still, Hill’s legacy is matched by few other reformers in its durability and popularity. It is instructive to examine what he proposed, and what happened.

To appreciate the complexity of the pre-1840 postal systems, we should note that

- in most towns, addressees had to collect letters at post offices,
- local delivery, in those towns that offered it, incurred additional charges,
- postal charges were paid by recipients, and local mailmen (where local delivery existed) often had to come several times to collect the fees,
- letters were charged by the number of sheets in them,
- distance charges were based on actual road distances, so letters between two nearby towns that had to go through London often faced huge fees.

To keep track of charges, and to prevent abuses by employees, detailed and voluminous records had to be kept by postal officials. (This feature, and the complexities listed above, were not peculiar to the British system. For a description of the labor-intensive record keeping in the U.S. Post Office of that era, see p. xxx of [John1].) High postal charges were a frequent source of complaint. So was the uncertainty introduced by the receiver-pays policy. Recipients could seldom be sure whether a letter would be worth the postal fee they were charged.

Sir Walter Scott liked to tell the story of how he had once had to pay “five pounds odd” in order to receive a package from a young New York lady he had never met: an atrocious play called *The Cherokee Lovers*, accompanied by a request to read it, correct it, write a prologue, and secure a producer. Two weeks later another large package arrived for which he was charged a similar amount. “Conceive my horror,” he told his friend Lord Melville, “when out jumped the same identical tragedy of *The Cherokee Lovers*, with a second epistle from the authoress, stating that, as the winds had been boisterous, she feared the vessel entrusted with her former communication might have foundered, and therefore judged it prudent to forward a duplicate.”

[Philonoe], based on account on
pp. 195-196 of [Lockhart]

At least the mass marketers who flood our mailboxes with junk mail do not ask us to pay for it! Not only did Scott have to pay, but he had to pay a fortune. In the early 19th century, five pounds sterling was about two months’ wages for a worker. Scott could afford these fees, but many people had to sacrifice food or clothing in order to pay for mail [Hill].

The high charges led inventive minds to numerous evasions of the postal monopoly. Some were crude abuses of the frank, sending private letters in the guise of official correspondence. Others were more clever. Since addressees were the ones who paid, they could inspect a letter and refuse to take it. Hence it was possible to transmit low-bandwidth messages for free, through deliberate misspellings in addresses. There was also a large (and illicit) business of smuggling letters. Some were carried by postal employees as a profitable sideline. Others were combined into packages and sent as such. Multiple letters (containing several messages to different correspondents in the same town on a single sheet, sent to one of them with instructions to distribute the other messages appropriately) were common, although they had been illegal since 1719. Many messages were written in newspapers, since those traveled for free.

A team of inspectors was employed to prevent the practice, but their task was hopeless. As one Inspector explained: “There are about 40,000 newspapers sent from London every night by the mails; of that number, with our present force, we are able only to examine from 500 to 1,000 each night... One person is only able to examine fifty in an hour.”

p. 145 of [Gregory]

The 1840 reform reduced rates, and apparently essentially eliminated the incentives for smuggling and other illicit maneuvers. This reform also greatly reduced the complexity and uncertainty by introducing a system of uniform charges paid by the sender. Any number of sheets could be sent for the basic penny rate, as long as the weight of the letter did not exceed half an ounce. The laborious computation and collection of proper charges by postal clerks were eliminated. The mail stamp (which gave rise to the hobby of stamp collecting and its supporting industry) and the home mail box into which letters could be deposited by the carrier were introduced at that time as well. Later mailboxes into which stamped letters could be dropped for pickup by the mailman were brought in as well. The reform simplified life for both postal employees and postal patrons, who used to have to bring their letters in person to a post office during official hours.

Popular historical accounts typically credit Hill with pointing out the huge cost of computing and collecting postal charges, and the small costs of transport. Hill calculated, for example, that the transport of a letter from London to Edinburgh by coach cost one thirty-sixth of a penny, as compared to the total postal fee for that letter of more than 500 times as much. Hill’s proposal of 1837 and the reform of 1840 are then often presented as simple victories of common sense over the benighted past. They seem like the obvious steps, once one understands the economics of the postal system as explained by Hill. The success of the reform then appears natural.

The true history is considerably more complicated. In the first place, the success of the reform even in stimulating usage was probably not as great as it might seem. The doubling in volume of letters in the first year of the reform may have come less from new correspondence, and more from abandonment of smuggling and other violations of the postal statutes. The savings from eliminating computation and collection of charges were not as large as Hill had promised. Hill had predicted that with a uniform prepaid rate, operations would be simplified to the extent that a four-fold increase in correspondence could be accommodated without any increase in costs. Yet costs (easily computed by subtracting the profit from the revenue figure in Table 12.1) went up slightly even though letter traffic went up only by 124% from 1839 to 1840. (Total traffic increased much less, since newspaper circulation through

the mails did not jump as much, and newspapers were the overwhelming majority of traffic by weight.) Furthermore, as will be discussed later, the fiscal effect of the reform was disastrous.

What is seldom realized, and apparently was not made clear by the public debates prior to the reform, is that Hill's proposal represented a huge change in the operating philosophy of the British Post Office. While the low costs of transporting the mail were not known to the general public, they appear to have been familiar to postal officials. This is evidenced by comments they made in response to Hill's proposal (see p. xxx of [Daunton], for example). The basic, although not publicly stated, philosophy was to charge according to willingness to pay. The reason for rates that increased steeply with distance was not that the costs of carrying letters a large distance were much higher. Instead, it was judged that the service of connecting people far apart was more valuable. (See Section 6 for more details and quotes. There is a wonderfully illustrative graph on p. xxx of [xxx], reproduced on p. xxx of [xxx], which shows that as postage fees were being raised and profit margins increasing in the 18th and early 19th centuries, the distance dependence increased sharply. The rates at the beginning of the 18th century were not uniform, but almost so, with just xxx rates.) The introduction of mail coaches in 1784, which led to a reduction in transport costs, was an occasion for raising of postal rates, on the grounds that the faster service was more valuable. Similarly, limiting basic rates to "single letters" had no justification on a cost basis, given the low cost of transport. It was assumed that people with longer messages would be willing to pay more. Those unwilling or unable to pay more, could and did resort to various data compression devices. One of them was the infamous "crossed letter," in which the first part of a message would be written in the regular way, and then the sheet would be turned sideways, and writing would continue right over the first part, perpendicular to it.

The British Post Office before 1840 was primarily maximizing profits. It is possible that they were good at that job. Experience with different rate schedules over the preceding century probably gave them a good understanding of demand at various prices. Certainly the reform of 1840 produced a huge drop in profits, a much more dramatic decrease than predicted by Hill. The Penny Post did not attain the profit levels of the old mail system of 1839 until a quarter century later.

Hill's reform changed the basic philosophy of the postal service to one of public service, based on notions of fairness. Hill is seldom regarded as a great economist, but his legacy is an enduring one. (For a fuller evaluation of his impact on economics, see [Coase2, Coase3, Daunton].) In particular, the uniform pricing he advocated has become a basic principle of public utility economics. We have already noted that pricing of postal services according to distance is already a considerable simplification of the optimal prices, whether those optimal prices are designed to match costs, or to extract

maximal willingness to pay. Hill accepted the argument that postal services should yield a profit for the government. His argument was that since the marginal costs of transport were extremely low compared to the price, the fair thing to do was to charge a uniform price for all letters. This would produce almost a uniform tax on each letter.

To come: further discussion of Hill's reform. His proposal was not for the Penny Post as implemented, but a two-stage system.

To come: discussion of U.S. postal system, operated on a different philosophy from the beginning of the Republic.

To come: contrast of USPS with Federal Express, which has implemented distance-sensitive rates (but is in a different category, handling relatively expensive and infrequent transactions)

Table 12.3. U. S. Postal Service rates for first class mail

year		price	hours of work
1799:	single letters		
	no more than 40 miles	\$0.08	0.8
	41-90 miles	0.10	1.0
	91-150 miles	0.125	1.25
	151-300 miles	0.17	1.7
	301-500 miles	0.20	2.0
	over 500 miles	0.25	2.5
1845:	single letters		
	no more than 300 miles	0.05	0.3
	over 300 miles	0.10	0.6
1863:	first half-ounce	0.03	0.2
1885:	first ounce	0.02	0.1
1999:	first ounce	0.33	0.02

13. Telegraph

Effective communication speeds improved with the growth of postal systems. There is a general perception that communication services were stagnant for centuries. They are reinforced by some data, such as the famous maps on pp. 426-427 in [Braudel], which show how long it took news from various cities in Europe to reach Venice. Between 1500 and 1750, these maps show there was little change, with messages from London taking about three weeks, and those from Lisbon about six weeks. However,

what such maps conceal is a likely increase in volume of communication, which led to faster effective communication. For affairs of state, such as news of an outbreak of war, where special messengers would be employed, it was the time to carry the news from one place to another that mattered. However, for general business or social correspondence, what mattered more was the effective speed of communication, which depends on the time between the writing of a letter and its delivery. This depended often on the frequency of the service more than on the speed of travel. It also depended on the distance from the nearest post offices, since senders and recipients had to go there to send and receive letters. In early times, these factors imposed substantial delays in addition to those of mail transport. John Hull's letter of 1671, cited in Section 3, waited over a month for its travel to start, a time comparable to the time it took the ship to reach its destination. Even on land, infrequent service meant that messages took a long time to arrive. In England, the dramatic growth in volume of mail in the 18th century, implied by the growth in revenues of the postal service (Table 12.1), brought with it increase in the number of post offices as well as more frequent service. That increased effective communication speeds by itself. Later, transport speeds started to increase, as road technology improved (even before the arrival of the railroads, see [Lay]) and general economic growth provided the resources and the demand for better transport. In England by the 1830s, mail coaches on the main roads were running twice as fast as in the 1780s [Gregory]. In the U.S., improvements were even more dramatic, since the country was less developed to start with. For details, see [Pred]. Speeds improved even further with the use of railroads for transport.

Interestingly enough, mail transport by rail initially cost more than by mail coach, both in the U.S. and in the U.K. This is an interesting precedent that helps put into perspective the observation of [Odlzko7] that in the late 1990s, most corporations were spending more to send data over their private line networks than they would have using a modem over the public voice networks. In both cases it was not cost savings that were the drivers for the adoption of the new technologies, but other advantages, primarily speed, flexibility, and ability to handle large volumes.

The improvements of postal systems in the early 19th century met much of the need for better communication that developing economies generated. However, mail was inherently limited to the speed of the fastest transportation technologies. Courier services, of which the Pony Express was the most famous, even though it was short-lived, offered some improvements, but they were limited as well to the speed at which people could travel. This spurred the continuing search for speed, for better methods to "annihilate both time and space."

Homing pigeons had been employed for particularly urgent transmissions for thousands of years.

Their use grew in the early 19th century. “With couriers and pigeons, Charles-Louis Havas [and his eponymous news agency] got news from the Brussels morning papers to Paris by noon and to London by 3 P.M. ... By 1846 there were 25,000 homing pigeons in Antwerp alone” (p. 12 in [Hearick]). Yet even pigeons did not satisfy the craving for reliable and near-instantaneous communication.

The most successful fast techniques for a long time were based on visual signaling. Some of these methods go back to antiquity. (See [Aschoff] for a thorough treatment.) The first widely deployed optical telegraph system was devised by Claude Chappe in France in the 1790s. The wars of the Revolutionary Period and of the Napoleonic Empire persuaded governments to lavish resources on fast communication. For the history of the Chappe system, as well as the ones it inspired in England, Sweden, the U.S., as well as in other countries, see [HolzmannP, Wilson2]. Some of these systems, such as the British Admiralty network connecting the main naval bases to London, were abandoned or scaled back once the Napoleonic Wars came to an end. Others survived much longer, and were even expanded. The French system was remarkably extensive, with over 3,000 miles of lines in 1844. Many of the French lines continued operating until 1853. A few Swedish lines continued in service up to 1881. I do not devote much time to the optical telegraph lines, since almost all were run by governments for their own use. Hence there is practically no pricing history to discuss. There were attempts to build commercial optical telegraphs in France in the 1830s [HolzmannP]. The first few did not succeed financially, and later attempts were stopped by a law of 1837 making telegraph communications a government monopoly. (The French were notorious for reserving new technologies for government use. The electric telegraph in France was restricted that way for its first few years.) The only semi-commercial traffic allowed on the French system was distribution of lottery results. However, there was at least one case of an illicit use of the optical telegraph for financial speculation that lasted for over a year before being detected.

One can speculate on whether optical telegraphs would have been developed further as commercial services had the electric telegraph not been invented. There was steady growth in demand for fast communication, and in willingness to pay for it. The semaphore systems did provide high speed, well over 2,000 miles per hour over long distances (p. 89 of [HolzmannP]). However, such speeds could be achieved only under good daylight conditions. In bad weather, messages could take days to reach their destinations, and some links in England were unusable for large fractions of the year (pp. 90 and 197 in [HolzmannP]). Furthermore, for commercial purposes it is not just the speed of a single sign that matters, but the throughput as well, and that was low, two to three signs per minute. (In its heyday, the Chappe system had 92 or 94 distinct signs available for correspondence.) Hence communication was

fast, but of low bandwidth. It appears from the estimates in [HolzmannP] that the typical transmission rate of the Chappe optical telegraph when it was functioning well was in the range of 10 to 20 bits per minute. Thus the maximal potential throughput was low, which meant that to support itself in commercial service, the optical telegraph would have had to charge high rates. Total volume even on the Chappe system was low, far lower than a commercial service surely would have carried. For example, only a few hundred messages were exchanged between Paris and Bayonne per year (Fig. 2.15 on p. 75 of [HolzmannP]). Most of the optical telegraphs in countries such as the U.S. were used for transmission of a few brief messages over short distances. Reminders of such systems exist in names of familiar landmarks. For example, Telegraph Hill in San Francisco is named after a semaphore used for announcing the arrival of ships.

Although optical telegraphs had serious shortcomings, and eventually disappeared, they did exert an enormous influence. They showed that practical communication faster than transportation was feasible, and in particular how an entire communication system, with provision for faults, etc. could be operated. ([HolzmannP] discusses this well.) They inspired many of the inventors who worked on the electric telegraph. Their influence on everyday thinking is illustrated by the association of the word telegraph with speed.

Fast commercial communications on a large scale had to wait for the appearance of the electric telegraph. Some of the analogies and contrasts with the Internet have already been discussed in sections 3 and 4. The electric telegraph alleviated (although did not completely solve) the main problems of the semaphore systems, namely lack of reliability and high cost (caused by low bandwidth). Even in the early days, transmission speeds of xxx words per minute were routine, which meant a data rate of xxx bits per minute, almost two orders of magnitude higher than for the Chappe system. This was still nowhere near the bandwidth of the postal system, say, but the compensating advantage of speed was crucial. (See [Lu] for an informative and entertaining discussion of the tradeoffs between speed and bandwidth.)

The telegraph was a revolutionary technology. The mail did what people could do by sending messages through friends or business travelers. It just did it more efficiently. The telegraph, on the other hand, was a dramatic innovation, for the first time providing a method of communication that was much faster than transportation. It put the whole world within instantaneous communication. Although its start was slow, after a decade it gripped the imagination of mid-19th century to an extent that appears to exceed the impact of the Internet. People spoke frequently of the “annihilation of space and time” that the new invention produced. It certainly had a dramatic influence on science and technology,

stimulating research on electromagnetism that led to many of key inventions that gave us the world we know. Many of the key figures in those areas got started in telegraphy. Lord Kelvin first attained fame for his analysis of the failure of the 1858 transatlantic cable, and Thomas Edison started out as a telegraph operator, and then a telegraph inventor. Even Alexander Graham Bell was working on improving the telegraph when he invented the telephone.

The world had always been more interconnected economically than is commonly realized. The Sumerians had traded with the Indus Valley, tin from Cornwall was used in the Middle East, and much of the silver that the Spanish colonies in the Americas produced was sent to China via the Manila Galleon in exchange for porcelain and silk. Yet the telegraph introduced an entire new dimension. The news of the Indian Rebellion of 1857 took 40 days to reach London. (It would have taken even longer were it not for the telegraphs in India and Europe that speeded up transmission.) Yet in less than two decades, news of a flood in India would affect prices at the agricultural commodity exchanges in Chicago within hours.

While there were many analogies between the electric telegraph and the Internet, there is also a major difference that is not discussed much in [Standage], for example. The Internet can be regarded as a “digital nervous system” for the world, to borrow a phrase coined by Bill Gates, a system that circulates information throughout an organization. The telegraph filled some of the same role, but it was too expensive and too unwieldy to penetrate deeply. One might be tempted to call it s a “digital spinal cord,” a system for transmitting only the most important messages to a few key points. However, it was not even that, since its bandwidth was too low.

The rise of the telegraph was even more impressive than that of the Internet. The latter could develop by utilizing the existing infrastructure of the phone system. In contrast, the telegraph had to build its own physical facilities from scratch, and yet it developed very rapidly. It helped that the infrastructure could be built incrementally. It could not be too small, however, as it had to be on a scale significant enough to provide real advantages over other methods of communication. The first attempt at commercial operation of the electric telegraph in the U.S. was Samuel Morse’s line from Washington, D.C., to Baltimore. Its financial failure was probably caused by a combination of two factors. One was the novelty of the technology. The other was the lack of demand for fast transmission between Washington and Baltimore, which were close to each other and did not exchange much commercial traffic. However, soon thereafter more promising lines were opened up, centered on New York City, the commercial hub of the nation. The new lines were financially successful. This led to much greater investments that allowed the whole country to be wired very quickly. Eventually truly gigantic

investments were mobilized for large telegraph projects.

The rational or irrational exuberance, reminiscent of today's enthusiasm for the Internet, that investors showed for the telegraph is illustrated by submarine telegraphy. The technology was not well understood initially, which led to numerous failures. "Of 17,700 kilometers of [submarine] cables laid by 1861, only 4,800 worked, and the rest were lost" (p. 18 of [Headrick]). However, technology quickly improved, and reliable transmission became routine. There were 1,120 kilometers of submarine cables at the beginning of 1861. (See Table 4.1 of [Headrick]. The discrepancy of that table with the 4,800 kilometer figure cited above is not explained in that work, and might be caused by looking at the start and end of 1861.) That figure grew to 46,065 kilometers in 10 years, and 147,513 in 20 years.

The scale of the private investments that the telegraph attracted is illustrated by the transatlantic cables. The first successful one was completed in 1866, after more than a decade of efforts (including some spectacular failures, especially the disappointing cable of 1858). The cumulative cost of the project, including that of the previous unsuccessful attempts, was \$12 million. To put this sum into perspective, the U.S. Postal Office expenditures in 1860 were \$15 million, and the budgets of the U.S. War and Navy departments were around \$10 million each that year. Also, Alaska was purchased from Russia in 1867 for \$7.2 million. Thus, the mobilization of \$12 million in an economy that was only partially industrialized, and in which the financial sector was much less developed than today, was a huge undertaking. It compares in scope to raising several tens of billions of dollars today. (It should be said that much, and probably most, of the funding came from British investors, who were also among the main providers of capital for American railroads.)

Table 13.1. International telegraph rates from New York City (per word)

year	London	Tokyo
1866	\$10.00	-
1868	1.58	-
1880	0.50	\$7.50
1890	0.25	1.82
1901	0.25	1.00
1924	0.20	0.50
1950	0.19	0.27
1970	0.23	0.31

To pay for the huge investment involved in the transatlantic cable, prices had to be high. Initially they were astronomical, as Table 13.1 shows. (The \$10 initial price for transmission of a single word from New York City to London was more than a week's earnings for a laborer.) Still, the demand

was there, even higher than expected. The savings that fast communication could offer were also astronomical.

Table 13.2. Telegraph rates (for up to 10 text words) from New York City

year	Philadelphia	Chicago	San Francisco
1850	\$0.25	\$1.55	-
1866	0.25	1.85	\$7.45
1870	0.25	1.00	5.00
1883	0.15	0.50	1.50
1908	0.25	0.50	1.00
1919	0.30	0.60	1.20
1951	0.60	1.00	1.60
1960	1.10	1.45	1.90
1970	2.25	2.25	2.25

Tables 13.1 and 13.2 show the evolution of pricing of telegraph services. The 19th century and early years of the 20th century show a rapid decrease in prices. However, later years show an increase. This is one way that the telegraph deviated from the pattern of other communication services. Table 3.3 shows even more basic differences. The telegraph is the only major communication technology to fade away. (The price increases in its last stages were part of a vicious spiral. Increasing prices led to decreased demand, which led to more price increases, and so on. Businesses whose costs are largely fixed benefit tremendously from increases in scale, but have difficulty coping with declines in volume of operations.)

A key point in considering the history of the telegraph industry, as was already pointed out in Section 3, is that it was never large. The high cost and inconvenience kept the telegraph from ever rivaling the mail, for example. Tables 3.1 and 3.3 show that even in 1890, there were over 70 pieces of mail delivered in the U.S. for each telegram, and this ratio kept growing afterwards. Revenue comparison is more favorable to the telegraph, but even there, the Post Office was always attracting at least three times as much spending as the telegraph.

Although the telegraph did eventually fade out, it was indispensable before the telephone became widely available. However, it was some time before it was learned how to use it efficiently. History records tragi-comedies such as the infamous encrypted telegram of November 1866 from the U.S. Secretary of State in Washington to the U.S. Ambassador in Paris [Weber]. The American Civil War had just ended, and the U.S. was pressuring the French government to remove their troops that were supporting the Emperor Maximilian of Mexico. That particular cable was part of the diplomatic maneu-

verings involved in this affair. The cost was almost \$20,000. This was an astronomical sum for those days, as it amounted to the annual wage bill for more than xxx workers. The entire budget of the U.S. Department of State for all foreign facilities, such as embassies and consulates, was under \$150,000 per year. Unfortunately, this enormously expensive telegram was completely and needlessly wasted. The message was encrypted to keep the British and French governments, through whose territory the telegram was transmitted, from reading the dispatch. Yet it is questionable whether this encryption served its purpose, since it was based on a weak cipher more than half a century old. Further, the message appears to have been intended to influence U.S. Congress more than anything else, and was promptly leaked to the press. The cryptosystem was also exceptionally poorly chosen for telegraphic communication. It turned a message into numerals, each of which was charged as a separate word. Still, nobody noticed anything amiss, and the telegram was sent. (The transmission required about six hours.) The unanticipated bill came later, and led to a dispute between the Department of State and the telegraph company that it took several years and the U.S. Supreme Court to settle. (The government had to pay the bill, with interest.)

With time, incidents such as that of the \$20,000 cable became rare. Even before this cable was sent in 1866, code books were getting printed, some for general commercial use, others for internal use of large corporations. They provided some confidentiality of messages, but the main purpose was to compress those messages to decrease transmission costs. For example, in *Harvey's Mining Code* of 1889, "archons" stood for "send two good Cornish miners," while in the general-purpose *ABC Code*, "pascoela" denoted "natives have plundered everything from the wreck."

Though the cost of writing a codebook was very high, the economies it provided were phenomenal. In social code, each code word replaced an average of 5.95 plain words, and in one commercial code, each code word did the work of 27.93 plain words. On the longer cables, where nine-tenths of the telegrams were commercial, 95 percent were in code. Thus between 1885 and 1898, while trade between Britain and India increased several times over, the number of words transmitted remained static as the increased use of codes gave the traffic an ever greater density of meaning.

p. 45 of [Headrick]

Telegraph rates were based on the number of words in a message, and therefore much hinged on the definition of a word. (There were also weighty issues concerning addresses. Should they be counted as part of the message, and if not, what forms would be allowed?) An arms race developed between the

telegraph companies, which would impose new rules, and codebook writers, who would rush to exploit the loopholes in the new rules. A brief description of this contest is presented in [Kahn, Standage], but the full history of telegraphic code books remains to be written. (For more concrete technical material, including some descriptions of the lengthy international negotiations, see [Friedman]. For a brief introduction, together with a description of a project to assemble a comprehensive bibliography of commercial code books, see [Reeds].) Nice mathematical techniques were developed for detecting and correcting errors in transmission, forerunners of the modern theory of error-correcting codes. In the end, simplicity won, when at the 1932 Madrid conference, the ITU decided to define a word as an arbitrary string of five letters (with the change effective in 1934). Telegraphic code books continued to be printed, though, as data compression continued to offer large savings. At least one corporate code book was produced as late as 1958.

Simplicity in rules about counting length of messages did prevail in the end. Pricing was a different matter. There was some development of differentiated services. The early telegraph rates were uniform (except for special deals for newspapers on some lines), as almost all traffic was generated by businesses [Flichy, Kieve]. Various experiments with pricing were attempted (such as a 50% surcharge for evening telegrams). Eventually services and pricing aimed at personal use were developed, such as special rates for telegrams to be delivered the following day. In the U.S., prices continued to be complicated well into the 20th century. A 1911 Western Union booklet of telegraph rates from New York City is almost 50 pages long. The rate for a day telegram to Buffalo, New York, for ten words or less, “address and one signature free,” was \$0.35, while the night rate was \$0.25. To Seattle, the corresponding rate was \$1.00, day or night. A “day letter” of up to 50 words, to be transmitted and delivered when there was slack time, cost \$1.50 to Seattle, while a “night letter,” also up to 50 words, to be delivered the next day, went for \$1.00.

Given how expensive the telegraph was, and how infrequently it was used, it should not be a surprise that it was not noted for simplicity. Still, telegraph rates did tend to become simpler with time. Not only were the rules about what is a word simplified, but distance dependence of rates decreased, or was eliminated entirely. This last development is illustrated in tables 13.1 and 13.2. As with other communication technologies, one can object that the decrease in distance dependence may not have had anything to do with user desires for simplicity. It could have been the result of a changing cost structure or of changes in elasticities in demand. However, the history of the telegraph does provide some convincing evidence that customer preferences played a major role in this area. In the U.K., telegraph service in the late 1850s was effectively a duopoly of the Electric & International Telegraph

Company (usually referred to as the “Electric”) and the British & Irish Magnetic Telegraph Company (the “Magnetic”). Both had rates that depended heavily on distance. A new entrant, the United Kingdom Telegraph Company (“UKTC”), managed to establish itself as a serious competitor by introducing a uniform rate of one shilling (\$0.25) for any telegram of up to 20 words. Both the low rate and the uniformity of the charge appeared to be major attractions (see [Kieve] for more details). The Electric and Magnetic companies were forced to match the UKTC rates on routes where they competed. However, in 1865, the Electric, the Magnetic, and the UKTC agreed on new rates that depended on distance. (This was before the days of serious anti-trust enforcement, when such collusion was legal.) For messages up to 20 words, the price was half a shilling within London and other large towns, one shilling for distances up to 100 miles, one and a half shillings up to 200 miles, and two shillings up to 300 miles. Thus although the exact length of the message did not affect the rate (most messages were under 20 words), distance did. The new rates were not popular (although it was probably the higher prices that mattered more than the distance dependent tariffs). When nationalization of the British telegraph companies was proposed, the promise of a uniform rate appeared to be a major attraction.

Uniformity of rate irrespective of distance was an integral part of the Post Office scheme. This system operated in Denmark, Netherlands, Belgium and Switzerland and also in the UK postal service, and it enabled the use of prestamped paper, which was especially attractive to the Post Office. This simplification, in the views of the officials, would make the telegraph more popular. Under the old system it was difficult for the public to know what the cost would be because the system of charges was so complicated. The basic charges of the companies related only to their own systems but not beyond them. So extra charges were the rule rather than the exception, and apparently this extra charge was always high in proportion to the ordinary tariff.

p. 162 of [Kieve]

Once nationalization did take place in 1870, a distance-insensitive tariff was introduced, and remained in place until the end of the service over a century later. Rates were modified many times, and other changes were made (such as in the length of the message that was covered by the basic rate, in whether the address was charged for or not, and so on, see [Kieve] for more details). However, the principle of distance-insensitive tariff was not touched, except for some special discounted rates inside metropolitan areas.

Why was the uniform one-shilling rate abandoned by the British telegraph industry in 1865? “In

February 1866, the UKTC stated that the uniform [one-shilling] rate for 20 words had been unremunerative to the company 'after four years persistent trials'" (p. 67 of [Kieve]). However, it is hard to interpret such a statement, since the UKTC's business had not been remunerative before or after the change. (The UKTC was half the size of the Magnetic, which was half the size of the Electric. It was hard to survive with such a small market share. The largest payoff to UKTC's shareholders came when the telegraph companies were nationalized on generous terms. The UKTC stock price went up over 350% between January 1867 and July 1869, when the Parliament authorized the purchase.) For the Electric and the Magnetic companies, the period of low simple rates was one of growing prosperity. Between 1861 and 1866, their combined revenues grew 54.3%, from \$1.535 million to \$2.369 million, while their combined net profits grew 79.2%, from \$0.497 million to \$0.891 million (p. 67 of [Kieve]). The growth in traffic during the period of low and simple rates was far higher than before or after, as is shown in Table 13.3. (The very high growth rates in the early 1850s shown in that table come from the very early period of commercial telegraph service. This growth was largely stimulated by large reductions in rates induced by increases in competition [Kieve], and by expansion of areas served.) Thus at least on the surface it is hard to explain the behavior of the British telegraph companies other than through blindness to the opportunities for growth through low and simple prices. (See the next section's discussion of the early history of the Bell System for a similar phenomenon.) A touch of the dot-com fever that inspires investments on blind faith in the promise of the future might have done wonders.

As with mail rates shown in tables 12.2 and 12.3, telegraph rates such as those adopted by the cartel of the Electric, the Magnetic, and the UKTC companies in 1865 already represent a concession to simplicity. They are based on distance alone. Yet willingness to pay is unlikely to be that simple to describe. Costs certainly are not. As with the mail, costs were lowest on routes with heavy traffic, primarily those between large cities. Thus the logic of socially optimal pricing on the basis of marginal costs would have led to a variety of rates specific for different routes. The same outcome would result from pricing on the basis of perceived value. Thus the Western Union rates, with a 50-page tariff table for New York City, were economically rational. Yet such rates tended to disappear. In the early days of telegraphy, when prices were highest, there was even more of an approach to optimal pricing. Towns concerned about lack of telegraph service would provide funding for construction of lines, or else would guarantee a certain level of business [Kieve]. (This is reminiscent of modern cities or states being concerned about being left out of the new telecommunications networks, and installing their own fiber.)

Table 13.3. Traffic carried by the Electric & International Telegraph Company

year	number of messages (thousands)	increase from previous year
1851	99.2	
1852	211.1	112.81%
1853	345.8	16.41
1854	572.1	132.76
1855	745.3	30.27
1856	812.3	9.00
1857	881.3	8.49
1858	870.1	-1.26
1859	1,025.3	17.83
1860	1,117.4	8.98
1861	1,201.5	7.53
1862	1,534.6	27.72
1863	1,825.4	18.95
1864	2,356.4	29.09
1865	2,971.1	26.09
1866	3,150.1	6.03
1867	3,351.9	6.41
1868	3,755.3	12.04

After nationalization and the re-introduction of the uniform one-shilling price for domestic messages, British traffic started to grow rapidly again. The compound annual growth rate was 18% a year from 1870-1 to 1874-5 (p. 183 of [Kieve]). This was faster growth than other European countries experienced. Financially, however, nationalization did not fulfill the promise of its advocates. Profits were lower than expected, and eventually became deficits. The decline of the telegraph in the U.K. came much earlier than in the U.S., which effectively had a private monopoly with Western Union dominating the industry. There seemed to be a variety of contributing causes. One was the high price paid at nationalization. Another may have been that a private monopoly was more efficient than a government one (although there were varying opinions on this subject, cf. [Kieve]). In particular, there was substantial expansion of telegraph lines under government ownership, largely driven by public pressure to provide universal service. Finally, the longer distances in the U.S. likely created more demand for telegraph services than the Britain had.

The standard telegraph charges around the world were universally metered, based on length of message, distance to destination, and sometimes time of day. However, there were also flat rate telegraph services. Corporations leased lines for their own private use, with fixed monthly charges. In

1868, about 10% of the British telegraph companies' revenues came from leasing private lines (p. 71 of [Kieve]), with about 11% of that 10% accounted for just by the royal family. In the U.S. in 1914, private lines made up 11% of the wire mileage used for telegraph services by the public carriers [Gabel2]. (AT&T had the dominant share of this market. The second "T" in the name stood for the telegraph. Although the company did not offer telegraph service to the public, aside from a brief period when it controlled Western Union, it did lease private telegraph lines.) Even this figure understates the capacity paid through flat rates, since it ignores telegraph lines operated by railroads for their own use. The relationship between railroads and telegraph companies was a symbiotic one. Railroads depended for the safety and efficiency of their operations on the telegraph. On the other hand, telegraph companies often depended on the railroads for the rights-of-way. The railroad requirements for telegraph service appear to have been a key factor in the extended persistence of separate telegraph lines. (Without those requirements, it probably would have been more economical for telegraph companies to switch to leasing lines from the phone carriers.) The interactions of railroads and the telegraph is a subject that has not been covered in detail in the literature. Some discussion of the division of responsibility in building and maintenance of shared facilities, and the contracts that were used, is presented in [Jepsen, Reid].)

In conclusion, the telegraph was unusual among communication services in its high cost, in infrequent use, and in its eventual demise. It never attained the stature of a real infrastructure technology. However, its history does show the desire for simple pricing among consumers, and the effectiveness of such pricing in stimulating usage.

14. Wired voice phone

The telephone attained its present preeminent position among communication services relatively slowly. For example, even in the United States, it was not until the 1910s that the revenues of the phone industry exceeded those of the postal system. (A comparison of data in tables 3.1 and 3.4 shows that today, the phone industry is four times as large, measured in revenues.) In most countries it took much longer to reach that stage. The main reason was that the phone had to build its own infrastructure. It also had to compete with the telegraph, which was almost ubiquitous. In many countries, the telephone was also hindered by governments that were concerned about competition for their mail and telegraph services. The variety of predictions about the future of the telephone that were made in its early days are presented in [deSolaP2]. There were visionaries who forecast the growth of telephony into a universal communication service. Still, it was apparently much more common to view its future in a much more limited way. One of the most fateful predictions was that of Western Union. After a short period of

intense competition, it decided to stick to telegraphy, as more promising than telephony. From our current perspective, that decision seems ridiculous. At that time, though, the important role of the telegraph and the inadequacies of the telephone apparently made it seem reasonable.

Table 14.1. International telephone prices. Standard rate for a 3-minute call from New York City to London

year	current dollars	hours of work
1927	\$75.00	200
1928	45.00	120
1930	30.00	80
1936	21.00	56
1944	21.00	40
1945	12.00	20
1969	12.00	6
1970	9.60	5
1974	5.40	1.6
1980	4.80	0.9
1986	4.83	0.7
1991	3.32	0.3
1995	2.40	0.2
1999	0.30	0.02

The early telephone had many technical and economic limitations. They undoubtedly helped blind people to the potentials of this technology. Initially the phone was a point-to-point service. Switching was developed soon afterwards, but, as was mentioned in Section 4, it was awkward and suffered from diseconomies of scale. Perhaps the greatest early limitation of the telephone was its cost. It was an extremely expensive technology. In 1900, basic monthly service in New York City cost about \$20 per month. This was about half the average monthly household income. (The corresponding figure today would be about \$2,000 per month. How many people would pay that much for Internet access?) Other cities did not suffer from diseconomies of scale to the same extent as The Big Apple. Even so, prices tended to be high. Phone service was primarily for businesses and the rich. However, with time technology improved, and economies of scale started working towards lower costs. Tables 14.1 and 14.2 show the steady decreases in prices that have occurred in the U.S..

The high initial costs of the phone went hand in hand with price discrimination and restrictions on allowed usage. The Bell Telephone Association, formed in 1877 to exploit the Bell patents, offered the following terms initially:

The terms for leasing two telephones for social purposes, connecting a dwelling house with any other building, will be \$20 a year; for business purposes \$40 a year, payable semi-annually in advance, ...

quoted on p. 7 of [Stehman]

At that time there was no switching, not even by operators, so the telephone was a point-to-point communications device. Subscribers were responsible for construction and maintenance of the line between the two phones they leased. All that the Bell Telephone Association provided were the two phone sets at the ends, and regular repairs on them. Thus the price differential between social and business use was a case of simple price discrimination, surely based on the belief that businesses would find the service more valuable than private individuals. Eventually an enormous literature would be generated, justifying charging business users more than residential ones on the grounds of heavier use, or greater load during peak periods. Historically, though, the price differential was just a case of charging according to willingness to pay.

In most countries, phone service started out as a government monopoly, and it is only recently that privatizations have begun to change that situation. According to [ITU, Mitchellxxx], basic service charges have tended to be only modestly higher for business than for residential users, and in many cases have been the same. It is ironic (but consistent with the evolution of airline pricing under competition) that the competitive and unregulated capitalist marketplace in the U.S. of a century ago would lead in discriminating against business users!

Telephone companies, faced with diseconomies of scale, were concerned with limiting, and not encouraging, use. They were undoubtedly also looking at the revenue potential of charging for additional services. This occasionally led to public relations debacles.

In Britain in 1889, postal officials reprimanded a Leicester subscriber for using his phone to notify the fire brigade of a nearby conflagration. The fire was not on his premises, and his contract directed him to confine his telephone “to his own business and private affairs.” The Leicester Town Council, Chamber of Commerce, and Trade Protection Society all appealed to the postmaster-general, who ruled that the use of the telephone to convey intelligence of fires and riots would be permitted thenceforth.

p. 102 in [Marvin]

It took much longer before hotel guests were allowed to use the hotel phone.

The high costs and diseconomies of scale created a compelling argument for metered rates. The Bell System in the 1880s experimented with pricing and determined that indeed metered rates were more profitable in large cities than flat rates. A concerted attempt was therefore mounted to move to metered billing. Announcements of the switch away from flat rates were made through public notices such as the following one:

The unlimited use of the telephone leads to a vast amount of unnecessary occupation of the wires, and to much borrowing of telephones by parties who are not subscribers. Thus the telephone system is so encumbered with calls which are unnecessary, and largely illegitimate, that the service is greatly impaired, and subscribers, to whom prompt connection is essential, become dissatisfied.

quoted in [NixG]

These moves away from flat rate plans met with intense resistance. For example, when metered plans replaced flat rate ones in Rochester, New York, in November 1886, almost all customers discontinued their subscriptions until AT&T offered flat rates again in May 1888 [NixG]. Still, per-call billing was widely used when the Bell patents expired in 1894.

The end of the Bell System monopoly on U.S. phone service in 1894 led to an interesting period of intense competition with independent phone companies. The effects of this competition on industry structure and market share have been known for a long time. However, it is only recently that the details of this struggle have been explored. The fascinating picture that has emerged is largely the work of the economic historians David Gabel, Kenneth Lipartito, and Milton Mueller [Gabel1, Gabel2, GabelW, Lipartito, Mueller, NixG].

The two decades from the onset of competition in 1894, to the Kingsbury agreement of 1913 that initiated federal regulation of the Bell System have many interesting aspects. Several are likely to be relevant in the Internet era. Lessons can be drawn from the competition between non-interconnecting rival that operated in the same neighborhoods, as well as from interconnection politics of that period. In this paper I concentrate on the extent to which customer desires for simplicity overcame sound economic logic. As was noted in Section 4 and earlier in this section, the arguments for price and quality differentiation, and especially for usage-based billing, were really compelling for the telephone systems of a century ago. The service was expensive, was used relatively infrequently, and had diseconomies of scale. Yet in face of all these factors, competition generally forced simplification of prices.

The independents' strategy in gaining entry was to offer lower prices and flat rate plans. This forced

AT&T to respond with similar moves. It lowered its prices, and it also to offer flat rate more widely. The story is told in more detail in [NixG]. That paper presents statistics showing that the proportion of Bell System customers who were on flat rate plans was positively correlated with the market share of the independents. We may owe the dominance of flat rate plans for residential calling in the U.S., and thus a large part of the success of the Internet, to this unusual period of competition.

The competition in telephony also led to a decrease in other forms of pricing that were economically rational. In the early days of telephony, basic rates often depended on the subscriber's distance from the central office. Such price differentiation could be justified on grounds of costs. (After all, it cost more to string the copper wire to a distant building, and maintenance costs were higher as well.) It could also be justified on the basis of greater value. (Somebody far from the central office was probably far from the center of town, and so had fewer alternatives, such as shouting across the street, or sending a child with a message down the street.) In any event, price discrimination based on distance was common, and was practiced by government monopolies as well as by private carriers. (The same story is being repeated in local data services, to be discussed in Section 18.) The interesting point is that competition tended to decrease this practice. For example, in Madison, Wisconsin, an independent competitor that was formed to challenge the established Bell company in the mid-1890s offered local rates independent of distance. This then led the Bell company to also switch to distance-insensitive rates (pp. 93-94 in [GabelI]). Today, such rates are the norm within large serving areas. Departures from distance independents in basic phone service do exist, but are concealed through carrier differentiation. (Rural areas tend to be served by independent carriers with their own rates. These rates are often artificially low because of explicit government-mandated subsidies from other carriers.) However, for any single phone company within a state in the U.S., rates are usually uniform, and the cross-subsidies this implies are ignored.

While competition led to a decrease in distance dependence of telephone rates, it led to an increase in the differential between business and residential prices. The 100% premium that was charged to business customers in 1877 had generally decreased over the next 15 years. The rise of the independent competitors starting in 1894 then led to a reversal in this trend, and an increase in the business rate premium. This rise was more pronounced in competitive markets than in monopoly ones (Table 1 in [GabelW]).

The discrimination in favor of residential users in pricing of basic phone service has persisted in the U.S. to this day, and became a part of the basic regulatory framework of universal service. (See [Mueller] for a history of the evolution of this concept, a history that is widely at variance with the

Table 14.2. Domestic U.S. telephone calling rates. Price of station-to-station, daytime, 3-minute phone call from New York City

year	Philadelphia	Chicago	San Francisco
1917	\$0.75	\$5.00	\$18.50
1926	0.60	3.40	11.30
1936	0.50	2.50	7.50
1946	0.45	1.55	2.50
1959	0.50	1.45	2.25
1970	0.50	1.05	1.35

common view.) Businesses typically pay twice as much as residential customers for a basic line, and then are charged for all calls, whereas residential users mostly still enjoy flat rate local service. There are no attempts to strictly police the distinction between private and business lines, even though there has been an explosive growth in the SOHO (small-office, home-office) business. However, regulators do attempt to charge such users more by imposing additional fixed monthly fees for multiple lines to the same dwelling.

The basic monthly fee for regular voice phone service does not depend on distance from the central office. Costs are averaged over wide areas.

Prices have declined, and their distance dependence has steadily decreased, and, starting in 1997, has been eliminated completely. There has also been no attempt to introduce quality differentiation domestically. When Sprint came out with its “hear a pin drop” ad, touting the quality of its fiber network, the other carriers rushed to install comparable technology. On the other hand, in some countries customers have at least at some points in the recent past had choice in overseas connections, with different carriers offering different degrees of voice compression, and different prices. This fits neatly into the general classification that inexpensive services tend towards a single high standard, and expensive ones are more likely to be subject to price and quality discrimination.

Long distance phone service in the U.S. shows an interesting pattern. Initially it was priced uniformly for everyone. The prices were high, as shown in tables 14.1 and 14.2, and traffic was light. For example, in 1930, the Bell System carried around 160,000 toll calls per business day, compared to about 300 million carried just by AT&T alone these days. The high early prices did depend heavily on distance, but were the same around the clock. The first time of day pricing was introduced in 1919, with three rates, the highest from 4:30 am to 8:30 pm, a lower rate from 8:30 pm to midnight, and the lowest from midnight to 4:30 am. I could not locate records of the considerations that led to the introduction

of rates, but the time bands argue strongly that the main motivation was not to shift demand and thus lower peak period traffic, but price discrimination, to allow private calls late at night at lower rates. Sunday discounts were only introduced in 1936.

As prices continued to fall, distance dependence decreased, but time of day discounts remained in place. Some forms of discrimination increased markedly, though; the subsidization of local service by long distance fees grew explosively in the 1950s, 1960s, and especially 1970s, as is shown in [Mueller].

The breakup of the Bell System led to substantial changes. The cross-subsidy of local service by long distance revenues started declining. Together with dramatic improvements in technology and competition, this led to further declines in prices. In the new competitive environment it also led to simplification of pricing. Distance dependence was eliminated completely, and time of day variation was reduced to typically at most two tiers. Some of the most popular plans, such as those from AT&T, have uniform rates around the clock.

So far I have discussed the evolution of the phone system in the U.S. alone. In most countries, phone service started out as a government monopoly, and it is only recently that privatizations have begun to change that situation. According to [Mitchell1], pricing of basic service has tended to be uniform, with business customers paying the same rates as residential ones. (Thus it is ironic that the competitive and unregulated capitalist marketplace in the U.S. around 1900 would lead in discriminating against business users!) There has been differentiation in pricing based on distance, but that has tended to decrease with time. Time of day variation in pricing (which is one way of discriminating in favor of residential users) came to Europe considerably later than to the U.S..

The greatest contrast between the U.S. and most other countries in pricing of phone service is in the approach to local calls. In the U.S. they are largely covered by flat rate fees, whereas in other countries they are metered. In the early days of the telephone, all local calls everywhere were covered by a fixed fee, since there was no technology to do the metering economically. Later, metering was introduced, initially charging just per call, and later, by the length of the call. In principle, the data should exist showing the effects of these transitions.

With advances in technology and reductions in access rates, we are clearly headed towards flat rate long distance service in the U.S. This was foreshadowed by the Sprint \$25/month unlimited weekend calling plan, announced in the fall of 1998.

15. Cell phones

The prominence of the Internet has tended to overshadow the rise of another great high-tech success, namely that of the wireless industry. There are more users of cell phones in the world than there are Internet users. They also produce much higher revenues than Internet users. Table 3.5 shows the growth of the U.S. wireless industry. Even in this country, with its high Internet penetration, cellular industry revenues are much higher than those from the Internet.

Cell phones are a nice counterexample to the claim that is sometimes heard that metered services cannot grow fast. Wireless services have been growing rapidly, but have been metered from the beginning, and rates remain high. The U.S. cell phone industry collects on average about \$0.25 per minute of wireless calls, as compared to \$0.10 per minute for the wired long distance phone service (and nothing for a large fraction of local calls). In data, the cost contrasts are even more striking. The estimates of [Odlyzko7] are that a megabyte (MB) of material downloaded from the Internet costs between \$0.02 and \$1.00 for various users, with residential modem users paying an effective price of about \$0.30. On the other hand, the initial pricing plan for Palm VII had a charge of \$200 per MB. (There was a monthly allotment that varied with the basic fee. It was 50 KB for the \$9.99 a month option, and additional transmissions were charged at \$200 per MB.)

The widely varying costs of wired and radio transmission are reflected in the universal use of compression for voice calls over cell phones. On the other hand, in land line transmission, compression had been used primarily on expensive international calls. (Although this is not widely known, there are many suppliers of reliable compression equipment for traditional telephony. Hence the bandwidth savings that are often cited as the great advantage of packetized voice are illusory.) In recent years, as costs of transmission have decreased, use of compression even on very long underwater links has decreased. The slight disadvantages in complexity, delay, and so on that compression introduces have more than compensated for the substantial bandwidth savings that it offers. This is a nice example of how simplicity has triumphed over efficiency. On the other hand, compression on the radio link reflects the scarcity of bandwidth in that environment.

Improvements in technology and greater spectrum allocations for communications will increase radio bandwidth. Still, this bandwidth is likely to remain orders of magnitude lower than that available on fiber optic backbones. Hence, even in the long run, although there may be plenty of bandwidth for voice calls, there will be a shortage for data, as mobile devices with powerful processors attempt to connect to big photonics pipes over relatively narrow radio channels.

In the intermediate run, the recent British radio spectrum auction is a bad omen for the predictions of broadband wireless communication. It brought in \$35 billion, far above early estimates. This sum means that the carriers involved in the bidding believe there is a scarcity of spectrum, in spite of the technological progress. It also means that they will not be tempted to price their services aggressively, since they will be trying to recover their investments.

Bandwidth fetches high prices, and in addition, as was discussed in Section 2, people are willing to pay far more for point-to-point communication than for broadcast “content.” This suggests that prospects for data communication may not be as rosy in the near term as some forecasts predict (and may also help explain why wireless data has not been a great success so far). The best prospects for wireless data are likely to be in narrowband messaging (email, chat). Multimedia “content” will likely not be a major factor for many years. On the other hand, it seems that the prospects are excellent for stimulating much greater voice calling, since cell phones are used far less than wired ones. Digital messaging can be very useful in that regard.

The scarcity of wireless bandwidth and its high price argue for quality and price differentiation. However, little of either is visible in the cellular industry. While compression plays a crucial role in wireless and there is continuing research on it, users have not been confronted with messages of the form “Dear subscriber, we are experiencing an unusually high load on our network, so if you wish your call to go through, it will be charged double the usual rate.” We also do not see premium services that would provide preferential connectivity to some customers. It is not certain whether this is due to the difficulty of implementing such solutions technically, or concerns about negative customer reactions (both factors that will matter on the Internet), but it is an empirical observation that carriers have not attempted to implement such schemes.

Although some time-of-day pricing plans have been introduced in cellular telephony, the general trend has been towards simple plans. Pre-paid plans, which limit users’ liability, are wildly popular. In the U.S., the most important innovation in pricing in recent years was the introduction of the AT&T Digital One-Rate™ plan in 1998. It was a form of block pricing, with initial rates of \$90 for 600 minutes per month, \$120 for 1,000 minutes, and \$150 for 1,400 minutes. One of its most distinguishing and most attractive features is that this price is all-inclusive, covering the long distance part of the call, as well as any roaming fees. The marginal per-minute revenue might appear to be only 7.5 cents per minute (as one moves from one block option to another), and roaming fees are often far higher than that. Thus this plan will definitely lose money for the carrier on some customers during some periods. However, it all averages out, since roaming is a fraction of total usage, and many customers use only a

fraction of the full allotment of time they pay for. Thus the carrier is playing the role of an insurance company, absorbing some losses but gaining from customers' willingness to pay for the protection of not having to worry about an accumulation of many small charges.

The Digital One-Rate™ plan has been widely imitated, and the current competition in the cellular marketplace in the U.S. can be characterized as a race for simplicity. There is complexity in the offerings, as different plans are tailored to different potential customers. In general, though, carriers are competing to offer plans that allow various family or employee groups to call each other at no charge, and so on. The effect on usage appears to have been dramatic. According to a press release from the U.S. cell industry association [CTIA], between the last quarter of 1998 and the last quarter of 1999, subscribers have increased their local calling from an average of 130 minutes per month in the last quarter of 1998 to 180 minutes per month in the last quarter of 1999.

The penetration of cellular phones in the U.S. is far lower than in Western Europe or Japan, and is especially low when compared to countries like Finland. The many incompatible systems present in the U.S. surely contribute to this development. In addition, there are other factors that appear important, factors related to the pricing structure. In the U.S., the cell phone owner pays for incoming calls. This imposes the risk of high charges not under the control of the user, and induces most people to keep their cell phones turned off except when they wish to make a call. The industry is trying to move to a "calling party pays" policy, but progress in that direction has been slow so far. Even if subscribers do not have to pay for incoming calls, other barriers will remain. In the U.S., initial purchases of handsets are usually heavily subsidized by the carriers. These subsidies are then recovered through higher per-minute charges. This may have been a good initial policy in the earliest days of the industry, when it was necessary to persuade people to try the service in the first place. Today, however, it might be better to raise the price of headsets, to move closer to a flat rate pricing system.

Not making a subscriber pay for incoming calls is likely to stimulate usage. Not making that user pay for outgoing calls is likely to be even more effective. It is surprising that the industry has not set up a system of 800-like numbers, in which companies would pay for the entire call from their customers with cell phones.

The popularity of prepaid calling cards also argues for calling plans that would be more flexible and yet still limit customer exposure. They might provide for fixed payments for blocks of time, and then require explicit authorization for the purchase of additional blocks.

Even with all the modifications suggested above, U.S. may have trouble catching up with the rest of the industrialized world in cell phone use. The problem is that cell phones will still have to compete

with free local calls. In Japan and Western Europe, cell phone calls are more expensive than wired calls, but apparently only by factors of two or three. In contrast, in the U.S. they are infinitely more expensive than the free local calls. (Recall the discussion of the dramatic effect of nonzero marginal rates, even low ones, in Section 9.) Furthermore, the flat rate structure in the U.S. has led to the high calling volumes shown in Table 1.1. Each wired phone is used for almost an hour each day in the U.S., far more than in other countries. Therefore replacing a wired phone with a cellular link is much more expensive than in Japan and Western Europe, at least if one wishes to maintain the same calling pattern. In Western Europe, where wired calling volumes are far lower, it is much easier to envisage a transition to a wireless environment. Thus, paradoxically, while flat rate local calling stimulates wired Internet access, it inhibits wireless calling. Hence the U.S. cell phone industry ought to be lobbying for imposition of metered local rates everywhere!

A dramatic demonstration of the relative merits of different services is provided by yet another comparison of cable TV to cell phones. In the U.S., the average monthly cable TV bill comes to about \$40, which is just about the same as the average monthly charge for a cell phone. The programming on the cable TV connection is watched by somebody in the household for about four hours a day, while a cell phone is used about 7 minutes per day. Thus the same fee brings 33 times as much usage with cable TV as with wireless telephony, when measured in the time spent with a service. Looked at from another perspective, people are willing to pay 33 times as much for a minute of cell phone conversation as for a minute of movies. The difference becomes even more dramatic when we consider the bandwidths of the signals. For both services, this is rather hard to define. However, one can make a reasonable case that a single cable TV channel transmits about 6 Mbps, while a digital cell phone receives compressed voice at about one thousandths of that rate. Thus even if we ignore all the other channels (usually in excess of 50) that the cable TV service provides but are not watched, we see that per bit, cell phones are valued by users 30,000 times as highly as television. One can quarrel with the exact numbers (for example, phone service provides a two-way channel, so the data rate should really be doubled, and so on). Still, no matter how the estimates are adjusted, cell phone bits are thousands of times more valuable than TV bits, as judged by people's willingness to pay.

The estimates about valuations of different uses for bandwidth suggest that the current preoccupation with content for wireless data services may be misplaced. Email, chat, location, and other low bandwidth messages can make cell phones much more valuable, and lead to more voice usage. However, it is voice that is likely to be the main revenue generator for most of the coming decade. It will be a while before we see streaming audio or video on cell phones. Technically it will be possible to

provide such services soon. However, that would mean restricting the number of users. It will be much more rational for carriers to sell that same bandwidth for voice calls. Furthermore, there are still great opportunities for expanding such usage of the spectrum. The number of users cannot grow much higher. On the other hand, the intensity of usage is still low. A cell phone in the U.S. is used only slightly more than a tenth as much as a wired connection. What this suggests is that there is great room for growth in wireless voice telephony. Such growth can be stimulated by both a modest amount of data traffic and pricing plans that come closer to customer preferences, as is discussed in Section 15.

The comparison of wireless telephony and cable TV above is not an apples-to-apples comparison, in that the great advantage of cell phones is the mobility they offer to users. However, it does reiterate the basic story told by Table 2.1 and the paragraph about postal services in the early 19th century, namely that content is not as valuable as point-to-point communication.

16. Minitel

France has twice deployed communication systems that were far more advanced than any others in the world. Unfortunately, in both cases better technologies were developed soon afterwards, and the French were slow in adopting them because of their commitment to the earlier indigenous systems. This occurred first with Chappe's optical telegraph in the 1790s, discussed briefly in Section 13. It predated the electric telegraph by half a century, and slowed the adoption of the new invention in France. A similar event occurred almost two centuries later, with Minitel. Designed in the early 1980s and deployed later that decade, for many years it made France the world leader in household use of online information and electronic commerce. It was far more successful than the teletext experiments in other countries. As recently as 1997, it was touted as superior to the Internet for ecommerce. Figures were cited that a greater volume of transactions was taking place over Minitel in France than over the Internet in the U.S. Yet this system turned out to be a dead end, and is now universally acknowledged to have retarded the spread of the Internet in France.

A complete analysis of the failure of Minitel remains to be written, but the report [OECD] serves as a good start. Many of the faults were basic ones, often caused by attempting to provide advanced features with a technology of the early 1980s that was simply inadequate. Other faults were caused by ignoring basic factors that are the main topics of this paper. Minitel was designed for database access and simple ecommerce. It was not designed human communications, although those did develop through chat lines. In particular, it had no email, and did not provide hard copies of transactions.

In addition to the problems listed in [OECD], it is worth emphasizing the terrible defect in pricing

Table 16.1. Usage of French Minitel system

year	hours of connect time (millions)	terminals (millions)	minutes of calls per terminal per day
1988	68.8		
1989	80.9	5.06	2.63
1990	91.5		
1991	97.1	6.00	2.66
1992	101.7		
1993	104.7	6.49	2.65
1994	102.5		
1995	110.4	6.40	2.84
1996	106.8		
1997	105.4	6.15	2.82

policies of Minitel. Charging was based on time spent in different areas, with some services charging upwards of \$100 per hour. Hence users had every incentive to minimize their interaction with the system. The result was limited usage. A comparison of Table 16.1 to tables 9.1 and 9.2 shows that Minitel subscribers spent one sixth as much time on the system as French Internet users do, and less than a twentieth the time that U.S. Internet users do. Naturally growth stopped early on.

The chief complaint about Minitel has always been about its cost [OECD]. This was likely a combination of high actual charges, and the psychological effect of metered rates. As we have seen in many context, usage-sensitive charging makes people much more aware of pricing, and reduces usage.

On a more global level, what accounts for the failure of Minitel? Some ascribe it to government sponsorship. Yet the Internet was also designed by government-funded researchers. Further, even the currently surviving U.S. corporate entries in the race to wire the world, namely CompuServe, Prodigy, AOL, Microsoft Network, and numerous others that have disappeared with hardly a trace (such as GENie), were failures in their original forms. They survived by transforming themselves into Internet access providers. They were able to do this more gracefully than Minitel primarily because they were more flexible, built on a PC platform as opposed to special purpose hardware. And that may be the main lesson from this story. Minitel was optimized to exploit the technology that was available at the time, and was optimized for purposes that its customers ended up not caring much for. The private online services were also optimized, with various proprietary interfaces and incompatible internal systems. They were also often designed for purposes that customers did not care enough about to make the services economic. (For example, AOL's origins are in a game service.) However, their underlying PC hardware and software platform was flexible enough so they could adopt, even though the adaptation

was painful (and many of their siblings did not make it). The system that triumphed has been the Internet, which was designed for flexibility, and not optimality.

17. Residential access to the Internet

For the mass public, the Internet became visible and accessible only in the mid-1990s. Even before then, though, there was a growing industry in online services, with CompuServe, Prodigy, and AOL the most prominent companies in it. Growth was rapid, but not at the rate we have witnessed on the Internet in the last half a dozen years. Each network had its proprietary user interface, and a limited selection of content providers. Communication among users was not emphasized. Pricing was based on a fixed monthly rate that covered a small number of hours of connect time, and fees for each additional hour. In many cases, some areas (for example of Prodigy) were designated as “premium” and charged at higher rates, and many of the databases on CompuServe had fees for items retrieved. In the early days, there were also charges for all email messages. All these fees were heartily disliked by the users. Email charges were dropped first. Then the industry converted to flat rate billing as a result of intense customer pressure (exemplified by the quote from [Swisher] in Section 8) and competition from new ISPs. (AT&T WorldNet was the first major ISP to offer flat rate pricing.) The dominant pricing model today in the U.S. is the flat rate, unlimited access plan for anywhere between \$14.95 and \$21.95 per month. Since most people have their local calls covered by a fixed monthly rate, their entire Internet access cost is flat rate. There are even some “free” access plans, supported by advertising.

The response of AOL’s customers to the introduction of flat rate pricing in October 1996 is shown in Table 9.1. Their usage tripled over the following year. This probably resulted from a combination of the new pricing scheme and the growing attraction of the Internet, which was going through a period of explosive growth stimulated by the wide introduction of the Web and browsers. (The increase in usage would surely have been much more rapid had there been capacity on the AOL network to accommodate the demand. As it is, AOL was for a while known to its exasperated users, who could not get online, as America Off-Line, and several lawsuits were brought against it.) Since that time, usage has continued to grow, but at a more moderate rate.

Most U.S. ISPs do have usage-sensitive billing plans. For example, AT&T WorldNet has a plan under which customers pay \$9.95 per month, and can stay online without further charges for up to 10 hours in that month. Each hour over those 10 costs \$0.99. The remarkable feature of such plans is that a large fraction of customers who pay for unlimited access would save if they switched to them. Average usage time online in the U.S. is about 25 hours per month (see Table 9.1 for AOL statistics,

which appear similar to those of other ISPs). Hence, even without examining detailed usage profiles, one can easily guess that many customers must be connected for under 20 hours a month. Even though they would save with the hourly plan, they choose to pay for unlimited access.

Another example of the preference for flat rates is that of AT&T's unlimited access plan. It was originally offered at \$19.95 per month. However, in the spring of 1998, it was replaced by the Standard Price Plan, which still cost \$19.95 per month, but covered only 150 hours per month of usage, with each additional hour charged at \$0.99 per hour. The justification for the change was that only about 5% of the users were connected for more than 150 hours per month, and that they accounted for an inordinately large fraction of modem time (the most expensive part in running an ISP). The intent was to discourage the small fraction of heavy users, and provide better service to the bulk of the subscribers. Yet in early 1999, a new plan was introduced, providing unlimited access at \$21.95 per month. While detailed statistics have not been released, it would not be unreasonable to conclude that many customers who do not spend anywhere close to 150 hours online per month were upset about the limitation on their freedom to do so, and are willing to pay the extra fee.

In the United States and Canada, local calls are predominantly covered by a fixed monthly charge, and so most people can access their ISP without paying for the time they spend connected. Hence their total cost is the fixed fee for the ISP account, and, in an increasing number of cases, a fixed monthly fee for the extra line they order for Internet usage. In most of the rest of the world, local calls are charged by the minute, with the rate typically depending on the time of day and day of the week. (This is what makes the "free" ISPs in Europe possible even without advertising; they do not charge any overt fees for Internet access, since they are paid through a fraction of the per-minute fees charged by the local phone companies to the users.) Thus even though ISP accounts typically carry fixed monthly fees, users pay by the minute. They do react to these prices. As the graphs in [Blair] show, usage jumps dramatically when low evening rates apply. However, the main effect of these charges is to reduce usage. AOL reports [xxx] that its users in the U.K. spend only a third as much time on line as those in the U.S. One could speculate that the difference in usage between the U.S. and Britain is cultural. Yet a recent study found that with flat rate pricing, British usage would climb to American levels [Durlacher].

The constraining power of usage sensitive pricing has led to numerous protests in Europe, and demands for introduction of flat rates for local calling. In the U.K., these demands have been reinforced by pressure from government regulators concerned about the lower penetration of the Internet than that in the U.S., and as of December 1999, BT has promised to introduce fixed fees, at least for Internet access. To come in next revision: details of British, German, and Japanese flat rate plans.

18. Non-Internet data networks

Most of the corporate spending for data transmission is for private lines, in which a dedicated connection of a fixed bandwidth is provided by telecommunications carriers between two points. (See [CoffmanO1] for estimates of the size of private line and public data networks.) Increasingly such links carry IP traffic, and are thus part of the Internet. Still, they are not visible to general users, since they are hidden behind firewalls. The price of such a link depends on the bandwidth, distance, and various regulatory and political constraints (especially for international lines), but not on the traffic carried. Most of the ISPs that do not own fiber networks carry their traffic over private lines leased from larger carriers. There is some discussion of private line pricing later, in Section 23. In general, as is shown in Fig. 23.1, and in a similar graph in [CoffmanO1] for longer private lines, distance dependence in pricing has been decreasing over the last two decades. In other respects, though, pricing of the long distance connections is becoming more complicated. In the days of strict government regulation, before the 1983 breakup of AT&T, it used to be simple, based just on the air distance between the endpoints of the link. This pricing was far divorced from reasonable cost allocation. It led large customers to place facilities in various out-of-the-way places, in order to minimize transmission costs for connecting their branches. This in turn led to intensive mathematical investigations of the Steiner tree problem and related questions. Today, in a competitive market, two conflicting tendencies can be discerned. On one hand, we are moving away from simple tariffs, and towards negotiations and spot market (as will be discussed in Section 23). This is especially noticeable for high capacity links. This trend is in accord with the general thesis of this paper, since prices of private lines, especially high bandwidth ones, are high. On the other hand, there is also an opposite tendency, in which corporations are outsourcing their entire data networks, and avoiding the complexity of dealing with a variety of carriers, price plans, as well as network administration itself.

While private line networks have been growing, their growth rates in recent years have been moderate, about 20% per year in bandwidth and 10% in revenues in the U.S., [CoffmanO1]. (Remarkably enough, these growth rates are similar to the 28% growth rate of data traffic estimated for the late 1970s and early 1980s in [deSolaPITH]. It might seem paradoxical that the growth rate would slow down in recent years, but that is probably accounted for by growth shifting to other data networks.) Much more rapid growth has been taking place in the Frame Relay and ATM networks, which have seen growth rates approaching and sometimes exceeding 100% per year. Frame Relay is expected to overtake private line in revenues, at least in the U.S., within a couple of years. Both Frame Relay

and ATM networks multiplex traffic from many customers. Most of the traffic is on PVCs (permanent virtual circuits), which are analogs of private lines in providing point-to-point connections. Pricing is predominantly on the basis of the size of the inlet to the public network, and independent of the traffic actually carried. (For the small fraction of traffic carried on SVCs, switched virtual circuits, there are usually usage sensitive charges, but for typical usage patterns, those are small compared to the fixed monthly charges.) Prices are also independent of distance, except that international virtual circuits are priced higher than domestic U.S. ones. (Generally there is no difference in pricing of international circuits among countries.) Thus the trend in this growth area is towards maximal simplicity.

In 1998, Sprint introduced its ION service, which allows businesses to combine voice, video and data traffic on one network. The initial announcement was that charges would be based on volume of data. However, it had to quickly retract this proposal in the face of vehement protests by customers.

19. Dedicated business connections to the Internet

The initial design decision about the Internet called for flat rate pricing. It was taken by ARPA. Dave Walden reports (personal communication) that the engineers at BBN working on the ARPANET were also discussing potential charging methods. They were in favor of the ARPA decision as a way to encourage growth in usage. These early decisions were the foundations for the ARPANET and subsequently the initial Internet billing approach of flat rates. The question now is whether the goal of encouraging use should continue to be pursued.

It is often claimed that all dedicated connections to the Internet in the U.S. are paid through flat rates. However, a substantial, although unknown, fraction are provided through so-called “burstable” rates that are illustrated in Table 19.1. Traffic is measured over 5-minute intervals, and the higher of the two traffic figures (for the two directions) is accumulated. At the end of the month, the top 5% of these 5-minute samples are discarded, and the highest remaining sample (the 95-th percentile) is used to determine the charges. Thus if in a 30-day month the lowest 8208 of the 8640 samples are below 384 Kbps, with at least one of those low 8208 samples above 256 Kbps, the charge for that month will be \$2500. These rates often appear attractive to users who have light traffic yet value the availability of broadband links to achieve low transaction latency, but have low average utilization. However, the pricing structure is such that in many cases users pay more with burstable than with flat rates. Unfortunately there are no studies on whether burstable rates are getting more popular or not, nor on what fraction of users who do pay burstable rates save with them over flat rate prices.

Burstable rates can be regarded as approximations to pricing according to the “effective bandwidth”

Table 19.1. UUNet burstable rates for Internet access:

flat rate T1 (1.54 Mbps): \$2500/month

burstable T1:

95-th percentile of bandwidth usage	monthly price
< 128 Kbps	\$1300
128-256	1900
256-384	2500
384-512	2750
> 512	3000

of a connection. The precise definition of this concept is technical, and beyond the scope of this paper. (See [Kelly], for example, for details.) Effective bandwidth measures in a quantitative way just how much burden a given traffic flow places on a network. Charging according to the effective bandwidth of a connection would have many desirable properties, but is impractical. Effective bandwidth is difficult to measure. Among other problems, it depends on all other flows. The 95-th percentile pricing of burstable rates is based on a very simple procedure that does capture some of the same effects as effective bandwidth. Still, the approximation is very crude.

Straightforward charging by the byte by service providers in the U.S. appears to be most common in Web-hosting, where a company contracts out the maintenance of its external Web server. Typically the customer pays a fixed monthly fee that depends on the bandwidth of the connection to the Internet and storage space. This fee covers some allotment of bytes that are sent out, and additional ones beyond that limit are charged at rates that are typically in the range of \$0.03-0.05 per MB.

Charging according to usage is also increasing as an internal corporate cost allocation mechanism in the U.S. With growth in data traffic and resulting expenses, central corporate networking groups are moving towards charging different business units for their services according to the volume of traffic those units generate.

Perhaps the most important conclusion one can draw from the frequent use of usage charging in the U.S. is that measuring traffic volumes is not too hard. It is often claimed that any kind of byte-counting would be extremely expensive and not worth undertaking, and would lead to endless billing disputes. However, neither of these problems seems to be a major obstacle, at least not for the simple byte-counting that is involved in either burstable rates or internal corporate cost allocations. (Attempting

to use pricing for real-time congestion control would be much more complicated, though.) Concerns have been raised about the fairness of charging users more just when service is worst (which is when packets dropped due to congestion cause retransmission), or for packets sent in error or by malicious outsiders. In practice, though, these problems appear not to be serious.

Charging by volume appears to be more common in other countries, especially those that have to pay for expensive links to the U.S., where much of the most popular Internet content resides. The book [McKnightB] has several papers and references to other papers that describe early experiments with usage-sensitive charging in places such as Chile and New Zealand. We next consider some recent examples, namely those of SWITCH in Switzerland, JANET in Britain, and Telstra in Australia. They all have some form of volume charging. Only a few details of their experience are available publicly, but we can try to deduce as much as possible from those tidbits.

As was noted in connection with garbage disposal pricing discussed in Section 11, usage-sensitive pricing is very logical, but it can induce undesirable behavior. This danger is especially acute on the Internet, since most of the intelligence is at the edges. Users can therefore interact with the network in much more imaginative ways than is possible with either garbage disposal or the voice phone system.

The SWITCH network provides connectivity to universities and research institutes in Switzerland. It is a non-profit organization that recovers its costs from member institutions. (Within institutions, these charges are not allocated to individuals or even to departments.) One third of the charges are fixed, depending just on the connection, and two thirds are volume-based [ReichLS]. Statistics on SWITCH traffic are presented in [CoffmanO2]. The link to the U.S. used to be saturated during business hours for several years, although its capacity was growing at a regular rate of over 100% per year. This appears to be the result of a conscious policy by SWITCH, developed in response to cost pressures. “[T]he costs for international connectivity exceed the costs for the national backbone” [ReichLS]. Links within Switzerland and to European countries are relatively uncongested. However, recently, perhaps as a result of decreased prices, capacity was increased enough to allow for much lower utilization.

What has been the effect of SWITCH pricing? There are both good and bad parts to the story, it appears.

The large part of usage-based charges has caused a high level of awareness of the costs of network usage within the organizations connected to SWITCH. Several creative methods have been devised by universities to reduce the amount of traffic and control the cost associated with their network connection. This includes reducing WWW bandwidth consumption by use of caching proxy servers (in some cases made mandatory through blocking of

direct connections), or converting USENET news feeds from “push” to “pull” mode.

Some of these measures to reduce traffic volume have actually been counterproductive in the sense that they reduced traffic only on those lines that were not congested anyway, so they did not actually reduce cost of operation. In some cases, efforts to save volume even made the service more expensive for everyone because of increased load on servers.

[ReichLS]

Unfortunately the paper quoted above does not contain quantitative estimates of the various effects of usage-sensitive charging. However, it goes on to say:

If the concept of volume-based charging is maintained in the future, it is likely that the prices will have to be further differentiated to encourage bandwidth savings in a way that is more focused to the expensive and congested or highly solicited parts of the network - currently the international links and in particular the connection to the U.S.

Now experience in both taxation and computer security strongly suggests that increasing the complexity of a broken system only creates more opportunities for clever folk to find ways around the rules.

Telstra, the dominant Australian ISP (and the former government telephone monopoly) has recently had problems persuading its residential customers of the virtues of usage-sensitive charging. However, the basic pricing scheme for business customers with dedicated links to the Internet has long been based on volume. They pay \$0.13 per MB of traffic received. (More precisely, they pay 19 Australian cents per MB, which is about U.S \$0.13 at current exchange rates.) There is also an option to pay \$0.10 per MB by opting for satellite delivery, which has much higher packet and transaction latency. In addition, there is a limited flat rate option, in which a fixed rate covers download traffic up to the 40% utilization level, with additional packets charged at \$0.19 per MB. This option is less expensive for certain utilization ranges that are high, but not too high. For details, see the full price schedule at <http://www.telstra.net>.

At least up to some time in 1998, Telstra used to have something called a Settlement Credit Plan that lowered fees for customers who generated Internet content. Such customers could get a credit of up to about \$0.03 per MB for traffic they sent out. This plan was terminated, presumably for abuses that one can imagine. Today, Telstra instead has what they call a backchannel tariff. If the traffic sent out by a customer is greater than the traffic received by a certain factor (2.75 for standard connection, 2.25 for the satellite option), Telstra levies a charge for that traffic as well. Thus even simple per-byte charges can quickly get complicated.

Table 19.2. Growth in Internet traffic. Shows terabytes sent from the U.S. to the British JANET academic network in January of each year (March for 1997).

year	TB
1997	3.73
1998	7.31
1999	16.08
2000	33.97

The final example of usage-sensitive charging we will consider is JANET, which provides Internet connectivity for British universities. It has also been faced with growing traffic and expensive links, especially those to the U.S. Starting in the fall of 1998, it began charging its member universities by volume, but only for packets coming from the U.S., and only between the hours of 6 in the morning until 1 the following morning. Table xxx shows the statistics on the total traffic coming from the U.S. It had been growing at about 100% per year before the charges were imposed, and it has continued growing at about 100% per year since. More interestingly, the fraction of traffic that is sent during the five free hours each morning has not changed appreciably. It has ranged from 8.2% to 12.7% since charging was introduced. That is about the same as in the few months before usage-based charges were imposed for which data is available.

We have seen so far that usage-sensitive charges can have problematic effects. Have they at least served to constrain usage? There is not much evidence of that. The attractions of the Internet have caused demand to rise, even with these charges. SWITCH traffic has grown at 100+% per year throughout the 1990s. Telstra does not release traffic statistics, but a collection of their network maps over the last few years (available on their Web site) shows that their trans-Pacific capacity has also been growing at about 100% per year.

20. Software agents and other technological palliatives

The histories of communication services presented in the preceding sections largely conform to the thesis of the Introduction; simple pricing along with higher quality, lower prices, and increased total spending is the natural evolutionary path for communication services. Deviations from this trend tend to be associated with expensive and infrequent transactions. However, the Internet is in many ways unique, and so historical analogies might not apply. In this and the next few sections we consider factors that might make historical analogies invalid.

As was mentioned several times earlier in the paper, yield management techniques are spreading to

a large extent because modern computing and communication technologies are making them possible. Manufacturers as well as service providers can get real-time feedback on sales and orders and tailor their production and pricing according to changing conditions. But we can turn this argument around. Similar computing and communication technologies are also available to consumers. Couldn't software agents be used to automate the price and quality negotiation tasks that have historically driven users to prefer simple pricing?

My prediction is that agents will be used, but only to a limited extent, and will not affect the drive for simplicity in pricing. In many high-tech areas, invocations of the words "software agent" or "genetic algorithms," or "fuzzy logic" are sometimes used to suggest that some technological magic will take care of all the hard problems without requiring serious thought. In practice, this has not happened. All these approaches have proved useful, but not one has been a panacea. There are two related factors that operate, and are likely to continue operating, in pricing as well as other fields. One is that these systems that are supposed to simplify life are not all that easy to master, and so are not as widespread as their proponents had hoped. The other is that even when these systems are used, they serve to encourage the growth of complexity that eats up any gain that had been achieved.

The main culprit behind the disappointing effects of seemingly helpful solutions has been the increasing complexity of our electronic environment. As an example, graphical user interfaces have made computing easier, but this has not lowered the degree of frustration that computer users experience. Gains in usability were lost through increases in the number and complexity of software packages [Odlyzko4, Odlyzko9]. More general discussion of the limitations of software agents is presented in [BrownD].

On an empirical level, we have plenty of evidence that online shopping will not usher a dramatically different world. Yes, there will be savings, and the economy will become more efficient, but not to the extent that the most ardent promoters claim. As an example, differences in prices among online booksellers are smaller than among brick and mortar ones, but they are considerable [BrynjolfssonS]. General inertia, and the proliferation of choices have negated the effects of greater price transparency. In general, even in the online environment, producers will continue to have incentives and means to erect barriers to "frictionless capitalism" [Odlyzko2].

In most cases sellers do not even have to create artificial barriers. Book buying online is now well developed, but few shoppers use any of the many free shopping robots to search for the best deals. Will this change, especially for more complicated situations? It takes effort to train a software agent, and can one train it well enough to cope with the increasing variety of choices? There are millions of VCRs

that continue blinking 12:00, because their owners did not feel it worthwhile learning how to set them. Will more people bother to train their software agents? Furthermore, even when they are trained, will those software agents simplify life? High-level executives have assistants who take care of scheduling their principals' time. Yet do these executives make fewer decisions than other employees?

Another way to deal with the complexity that optimization requires and at the same time provide the end users with the simplicity they desire is to have intermediaries. These enterprises would negotiate with communication providers for basic services (using all the modern tools of auctions, futures, and derivatives) and offer them in packages to users at flat rates. This proposal is feasible, but does not refute any of the arguments for simplicity presented here. It basically consists of renaming various agents in the communications marketplace. The evidence of this paper is that simplicity is of paramount importance for individuals and small organizations. The company that is responsible for their communications needs is the effective carrier, whether that company has its own physical facilities, or leases them from other carriers.

21. The role of distance distribution of traffic

There has been frequent talk about the “annihilation of space and time” at least since the early 19th century. (See the quote in Section 6, for example.) The electric telegraph was often claimed to achieve this ardently sought goal. However, the telegraph left much to be desired, as it was an expensive and awkward technology. The telephone was in general a substantial advance over the telegraph, and provided much more flexible communication. Now the Internet, especially with its wireless extensions, is claimed to finally achieve the “death of distance” [Cairncross1, Cairncross2].

The “death of distance” thesis is often pushed to excess, interpreted to mean that location won't matter, and that we will live wherever we wish, and telecommute to jobs halfway around the world. Such predictions are heard most often in Silicon Valley. Yet it is easy to find evidence contradicting this thesis. After all, Silicon Valley is the place that many startup companies feel compelled to move to. Silicon Valley companies are also the main advocates of increasing immigration of skilled workers. These companies choose to employ such engineers and scientists in California for ten times the wages at which they could be put to work in their native countries. As another example, Goldman Sachs, one of the most prominent Wall Street investment banks, has long had an office in San Francisco. However, in 1999 it opened a large new branch in Menlo Park because “the company's deal makers came to realize that [San Francisco] was too far (about 30 miles) from the Silicon Valley action” [Amer].

Clearly physical proximity continues to matter. While there is growth of telecommuting, and some

software jobs do get outsourced to India or Singapore, close personal contact is crucial for many jobs. As a result, business travel is booming in parallel with growth of teleconferencing, fax, voice phone, and the Internet. Even though stock and commodity exchanges are abandoning their old-fashioned trading floors, new trading floors of equal size are being built by individual investment banks. The employees at those floors sit at desks with numerous electronic communications devices, but the floors are sloped and desks arranged so that groups of traders can be in eye contact with each other.

All that “death of distance” really means is that communications prices are becoming independent of distance. Hence our decisions about social and business interactions can be made without regard to monetary costs of electronic communications. Those interactions will continue to be constrained by other factors that have traditionally produced a localization of our activities.

Mail, telegraph, and telephone traffic all show a strong distance dependence, with most traffic local. This phenomenon was already observed in the 19th century (see [Isard] for detailed discussion and references), although its discovery is often ascribed to Zipf [Zipf]. Usually a “gravity law” provides a good fit to the data, in which the intensity of interaction between two cities is proportional to the product of their populations divided by some power (usually between 1 and 2) of the distance between them. (The fit is seldom perfect, and often one can do better by taking into account more details of the geography of a region, as in [KowalskiW].) This distance dependence is not just an artifact of pricing, since it applies even to mail, where, as was noted before, prices have been independent of distance for over a century.

The Internet appears to be a striking exception to the locality of traffic. Currently it is commonly thought that most of Internet packets travel extremely long distances, although no careful quantitative studies are available. As one illustration, the network maps of Telstra, the dominant Australian ISP, show that most of its traffic appears to be with the U.S., as even the domestic Australian links seem designed to distribute U.S. traffic internally. According to some statements by Telstra officials, 60% of Telstra Internet traffic is with the U.S., and 85% is international [Taggart]. Unfortunately we do not have enough data to tell how typical this is. In contrast to the Australian experience, for example, one large Korean ISP has claimed that only about 10% of its traffic is exchanged with the U.S. Some European carriers report that their domestic traffic is growing faster than that with the U.S. On the other hand, for the British JANET academic and research network, transatlantic traffic in 1998 and 1999 grew faster than traffic through LINX, the London public Internet exchange that is apparently the largest one in Europe.

As was noted in [CoffmanO1], one of the key questions for the future of the Internet is the distance

distribution of its traffic. The optimal architectures of communication networks depend on answers to this question, as do prospects of some novel technologies. If traffic were to be largely local, photonic technologies that allow long distance transmission without regeneration would have limited markets. On the other hand, if a large fraction of traffic were to go halfway around the globe, such technologies would be vital in lowering costs.

There are conflicting tendencies on the Internet. On one hand, centralized databases simplify updates, maintenance, and load distribution, and many corporations are pursuing this strategy. On the other hand, locality of data may correspond better to the patterns of our other interactions, and it helps to avoid congestion on the Internet. (Note that even in the absence of congestion, there would continue to be incentives to store data locally because of speed of light limitations. A photon takes 200 milliseconds to travel from the East Coast of the United States to Australia and back. Hence if you are in a hospital in New York City, say, you would not want a surgeon in Sydney to operate on you, no matter if she was the world's best!) At this point we do not know which tendency will dominate. The imbalance in Internet traffic between the U.S. and the rest of the world has been increasing in the last few years (with more bytes leaving the U.S. than arriving), but there is also vigorous growth of local traffic, so we do not have the data to form a comprehensive picture. A major question is whether any special pricing moves are called for in order to change the distance distribution.

A change in the distance dependence of Internet traffic would not be unprecedented. Such changes have occurred in the past. North American mail in the 17th and 18th centuries was used largely for long distance communication, between the British colonies and England, and between merchants in distant cities within the continent [John1]. Eventually, local traffic grew to dominate. A similar trend towards greater locality of mail usage can be seen in the data for the Swiss postal system in Table 21.1.

The telephone in its early decades was used primarily for local communication. Probably the main reason that Western Union agreed to withdraw from phone competition with the Bell System in 1879 was that the phone was expected to remain primarily a local service. Hence Western Union was more concerned about possible competition from AT&T in the lucrative long distance telegraph business than about getting into the new-fangled phone business itself. Eventually, though, long distance phone service did grow. Today, interstate calls make up about 14% of the minutes on the U.S. phone network, intrastate toll calls another 11% percent, and local calls are down to about 75% [FCC1]. Thus it is possible that Internet traffic will also evolve.

One of the arguments for new Internet pricing models is to introduce distance sensitivity to the charges. The perception is that the flat rate pricing system provides cross-subsidies that artificially

Table 21.1. Increasing locality of Swiss mail

year	letters (millions)	letters within local region
1850	9.9	5%
1852	12.0	20
1855	14.4	19
1858	16.9	21
1861	19.0	20
1865	24.1	25
1870	30.8	25
1880	42.4	32
1890	57.4	35
1900	88.9	39
1910	139.0	37
1920	194.2	35
1930	228.9	38
1940	239.6	39
1948	356.5	39

subsidize wasteful long distance transport. A possible approach to overcome this is to follow the example of the JANET network (discussed in Section 19). It has imposed charges for transatlantic traffic only. This encourages local caching and discourages long distance traffic.

While one can use pricing to encourage local traffic on the Internet, the case for doing so is questionable. The reason is that long distance transport is not likely to be the dominant part of the cost structure, even if all traffic goes half way around the globe. Bandwidth costs scale nonlinearly, with larger pipes much less expensive per Mbps capacity than smaller ones, at least up to the limits of bandwidth that technology can provide. (See [CoffmanO1, FishburnO] for examples of pricing for various links, and discussion.) Hence it is possible to construct abstract economic models which show that for reasonable distributions of demands, even if traffic from all sources were distributed uniformly over the globe, the long distance part of the network would not dominate the costs. This is consistent with the estimates in [Odlyzko7] that the backbones of the Internet are relatively inexpensive to operate. It is also consistent with the general picture of the Internet as a complicated network in which most of the intelligence is at the edges, but so are most of the costs.

Abstract economic models are one thing, but actual costs are another, and ISPs have been complaining about the costs of long distance transport. Complaints have been especially vocal about transoceanic routes. However, even in this area, recent reductions in prices have made a large difference.

(See Section xxx for more detailed discussion.) For example, the British academic JANET network charges its member universities two British pence (approximately \$0.03) for each MB of traffic received from the U.S.. In the U.S., the typical residential customer with a 28.8 Kbps or even a 56 Kbps modem downloads about 2 MB per day. Thus at the rates charged by JANET, the incremental cost of carrying traffic across the Atlantic is only about \$2.00 per month, roughly 10% of the flat rate fee. For cable modem or DSL customers, who typically pay between \$40 and \$60 per month, and download between 10 and 20 MB per day, the incremental cost would be between \$10 and \$20 per month, which is much more substantial. However, the JANET rates are high compared to some recent prices for transatlantic capacity. For example, the spot market prices for OC3 (155 Mbps) links across the Atlantic that are cited in Section 23 come to about \$100,000 per month. Suppose we run such a link at an average utilization of just 20% in the more loaded direction. (JANET was running its two OC3 links from the U.S. about about 33% utilization in January 2000. A year earlier, its smaller T3 links were run at over 50% load. Thus a 20% utilization is very conservative, and would provide excellent service.) We then find that the incremental cost of traversing that link is just \$0.01 per MB. That lowers the costs for residential users cited above by a factor of three. The main fact to keep in mind is that the costs of running corporate private line networks were estimated at over \$0.50 per MB in [Odlyzko7]. Thus as long distance transport costs start declining more generally, it will be the costs at the endpoints, especially the costs of network administrators, that will dominate even more than they do now. We really are facing the “death of distance,” in which we do not have to worry whether our traffic is local or goes halfway around the world. However, traffic may very well be predominantly local for other reasons than costs of transport.

22. The tragedy of the commons

The Internet is often claimed to differ drastically from other communication services in being uniquely subject to “the tragedy of the commons.” (See, for example, [GuptaSW].) Users who pay flat rates supposedly have incentives to saturate their links. In voice telephony, there is a natural limit to usage, as people have limited time. No such limit is present on the Internet. Indeed, as has often been pointed out, every link has always had to be upgraded, as traffic growth has been inexorable. Thus it is argued that only usage sensitive pricing can restrain demand enough to provide high quality transmission.

The tragedy of the commons is a widespread phenomenon. Many fisheries have been ruined by overexploitation with individual boats having an incentive to catch as much as possible, irrespective of

what this might do for the future. Road transport also exhibits this phenomenon. New superhighways that are built to relieve congestion lead people to move further away (to get away from the noise of the central city, to allow a spouse to take a better job, ...), bringing congestion back. The time that people are willing to devote to travel each day is about constant, as is the fraction of their income that they spend on travel [Gibbs, SchaferV]. Thus reducing monetary costs of transportation induces more travel, and building faster links does the same. This dynamic applies to public transport just as much as to car travel, with people essentially commuting between London and Paris via the Chunnel train, or between Boston and Washington by air. There has been a continuous increase in distance traveled per person over the last few centuries, paralleling the growth of communications (although at a slower pace). Under such tragedy of the commons condition, especially when the costs of installing new capacity is high, measures to constrain demand are appealing (and some results are discussed in Section 11).

Table 22.1. Utilization of phone lines in the U.S.

year	lines (millions)	local calls (minutes per day per line)	local calls (minutes per day per person)
1980	102.2	39	17.5
1984	112.6	40	19.1
1988	127.1	39	20.2
1992	143.3	37	20.8
1996	166.3	40	25.1
1997	173.9	42	27.3

The tragedy of the commons argument does not apply to the Internet. This is demonstrated at length in studies such as those of [Odlyzko6, Odlyzko7, Odlyzko8], which I can only briefly summarize here. It is useful to first note, though, that there are numerous situations where the tragedy of the commons does not occur, even when the setting appears to invite it. For example, it was already noted in Section 11 that a noticeable fraction of water supply systems in the U.S. have flat rate billing. Yet demand for water has not exploded. Another example, perhaps more relevant for the Internet since it is in communication, is in local phone calling in the U.S. As was noted in [Odlyzko7, Odlyzko8], and is shown in Table 22.1, the volume of local calling (including fax and modem calls) per line has stayed remarkably constant over the last two decades. (It did increase substantially over the first few decades of telephony, but not in recent times.) On the other hand, local calling per person has been growing rapidly. This says that the value of a phone line is largely in its being idle most of the time, available for

brief bursts of usage on demand. This is a different dynamic than that of road transport, and prevents the tragedy of the commons from pushing up local calling per line.

A dynamic similar to that of local calling in the U.S. also operates in data transport. The basic reason that high capacity links are unlikely to be saturated is the same one that makes our PCs and our LANs idle most of the time. What users want from data networks is not so much to transmit many bits but instead to carry out transactions quickly. When they can afford it, they pay for big pipes in order to obtain low transaction latency, and leave those pipes empty most of the time. That is just the behavior we see with PCs, which feature fast processors just for the brief spurts when spreadsheets need to be recalculated, or word processors loaded. The fact that most corporate private line networks in the continental U.S. are lightly loaded is a refutation of the tragedy of the commons argument all by itself. To make the refutation even more convincing, one can look at the examples in [CoffmanO1, Odlyzko6, Odlyzko7, Odlyzko8] of links whose capacity increased by large factors without causing an increase in traffic beyond the usual growth rate.

The studies of [Odlyzko6, Odlyzko7, Odlyzko8] show the key role that prices of transmission play in determining how data links are used. Most of the Internet consists of private line networks [CoffmanO1], and inside the continental U.S. those are mostly lightly utilized. On the other hand, the most more expensive trans-oceanic private lines appear to be much more heavily used, and so are the links to Internet backbones from ISPs that aggregate residential modem traffic. Thus it appears safe to predict that when transmission prices start declining at rates approaching those that technological progress suggests are feasible, users will be able and willing to pay for uncongested links and although their traffic demands will grow, they will not have the incentives to saturate those links. However, for that to happen, prices do have to decline.

23. Data transmission prices

The prices of data transmission are a great anomaly. Contrary to the general perception, these prices have for the most part been rising in recent years. This is particularly surprising for those aware of the stunning progress in photonics, progress faster than that of Moore's Law in semiconductors. Just about all other products and services where demand is rising and technological progress is rapid have seen consistent and rapid decreases in prices. However, for data transmission, the history of prices for private lines in the U.S. is illustrated in Fig. 23.1. It shows the tariffed rate for a T1 (1.5 Mbps) link from New York City to Pittsburgh, a distance of about 300 air miles, which is about the average for long distance private line links. (For a similar price history, but for a longer connection, see [CoffmanO1].

In each case, the price shown is in current dollars. It ignores the costs of local connections, which can add up to \$1,000 per month to the cost. It also ignores the discounts for bulk purchases and long-term contracts, which can lower the effective price by up to 50%.) This graph shows that after declining steeply in the 1980s, prices bottomed out in 1992, and have since then climbed by over 50%. (An interesting illustration of the perils of forecasting transmission prices is presented by the paper [Irvin]. It was published in 1993, but was based on data through 1992, and predicted, with the help of two separate economic models, that by 1998 prices would decline by about 50%. Instead they rose by more than that.)

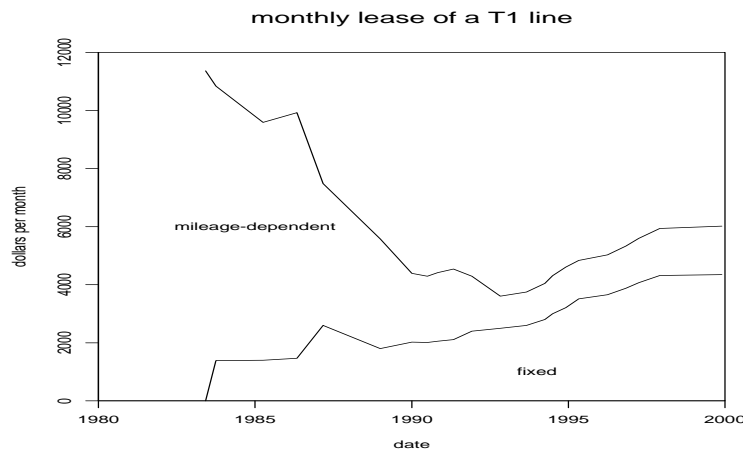


Figure 23.1. Tariffed prices for 1.5 Mbps private line from New York City to Pittsburgh.

The general expectation in high-tech pricing is of consistent and rapid decreases, such as we have seen in PCs. However, even the computer industry story is more complicated than is normally observed. The prices of various components of the PC have not been moving in lockstep. In microprocessors, the price of the latest chips been basically steady, while their performance has been increasing rapidly, and Intel revenues and profits have grown at fast rates, as is shown in Table 1 of [FishburnO], for example. (The low-cost PCs of recent years were made possible to a substantial extent by reliance on microprocessors that were far from the most powerful ones on the market.) In software, the price of the Microsoft operating system has been increasing, while functionality has been growing. The really interesting case is that of DRAMs. In the early 1990s, the price per bit of memory of DRAMs staid basically constant. Then, in 1995, a rapid decline started, which over the last four years has brought prices down to about the level that Moore's Law would have predicted for today. Microprocessors

and DRAMs both require large upfront development costs and hugely expensive fabrication facilities. However, in microprocessors, Intel's effective monopoly combined with a healthy regard for competition (as exemplified by Andy's Grove famous saying, and the title of his book, "only the paranoid survive") led to steady progress for consumers. In the competitive DRAM market, imbalances between supply and demand led to anomalously high prices for half a decade, and it is only recently that this imbalance has been corrected.

The data transmission market is even more complicated than that for DRAMs. Huge investments are required to build networks and typically two local and one long distance carrier are required for effective end-to-end service. The increase in U.S. private line prices in the 1990s was a response to increasing demands. Was it unjustified? One could easily argue that it was the low prices in 1992 that were unjustified. In a complicated system that serves many purposes, cost allocations are often unavoidably arbitrary. In the late 1980s and early 1990s, there were often concerns as to whether all the fiber that had been laid would ever be used for slowly growing telephony. Carriers may have been motivated to lease links at low rates in order to maximize revenue, without considering the sunk costs of installing that fiber. (One of Irvin's models [Irvin] was based on that assumption.)

Prices were driven up in the 1990s by unexpected increases in demand for data transmission. An extra boost to the demand came in the late 1990s from the Internet, which grew faster than traditional private line business. There are several reasons to expect rapid decreases in transmission prices in the near future. Technology is advancing and there are many competitive carriers. Perhaps most important, data transport is no longer a small sideline for the voice telephone networks. Whereas before, data transmission capacity could be priced either high or low, as division of costs between that area and the dominant voice network was largely arbitrary, in the future most of the network will be used for data. Hence trends in data transport prices are likely to track technological progress, at least when averaged over large areas and large time spans. One feature of competitive markets, such as that of air travel, is that there are many artificial constraints that lead to annoying irregularities in pricing.

A general decline in data transmission is beginning on heavily used routes. Some of the most dramatic and most visible decreases have taken place in the transoceanic markets, where prices were exceptionally high. In the mid-1990s, the tariffed price for a T3 (45 Mbps) line from New York to London was about \$1 million per month. In early January 2000, it was possible to lease an OC3 link (155 Mbps, or 3.4 times the bandwidth of a T3 link) for one fourteenth of that price. See Table 23.1 for some recent bandwidth quotes.

The emerging spot market for bandwidth is showing some interesting features, ones that are not

unexpected given the behavior of other competitive markets. Opportunities for arbitrage are opening up (with some recent offers for direct New York to San Francisco links higher than the combined New York to Los Angeles and Los Angeles to San Francisco route). There are now several companies, such as RateXchange, Band-X, and Interxion, which are creating a market in bandwidth. The spot market and bandwidth brokers are evolving slowly, and are still nowhere near real-time capability, with contracts requiring several months to take effect. There are some questions as to how far the bandwidth brokerage business will develop, since data transmission has a specificity of location that is not present in the corn or even electricity markets. In particular, since most of the costs in data transmission are local, the spot markets may not affect them much.

Table 23.1. RateXchange spot market for bandwidth, end of January 2000.
Prices for OC3 (155 Mbps) links, per month for 12-month leases.

connection	price
New York - Los Angeles	\$85,000
New York - London	70,000
New York - Paris	125,000
New York - Amsterdam	125,000
New York - Seattle	92,500
New York - Chicago	26,513
Los Angeles - Seattle	35,984
Los Angeles - Denver	30,876

In spite of the price increases in the 1990s, the U.S. has the lowest data transmission prices in the world. A recent European Commission report [EU] cited comparisons showing that for comparable distances and bandwidths, U.S. private lines are four times less expensive than ones within the same European country, and 16 times less expensive than international links within Europe. The Commission is now issuing directives to European carriers, forcing them to reduce their rates. The public policy has changed. It used to encourage (or at least tolerate) charging high rates to corporate users to subsidize residential phone subscribers. The current priority is to encourage ecommerce and Internet penetration.

Even without overt government intervention, prices in Europe have in many cases been decreasing. Competing carriers are emerging, and the market is growing. As an example, the TEN-155 research network (operating with 155 Mbps links) reportedly has about the same budget for transmission as its predecessor, the TEN-34 network (with 34 Mbps links) [SabatinoA].

In conclusion, prices of data transmission are likely to start a rapid decline soon. The outcome is likely to be the same as in other high-tech industries, namely increased demand and eventually also

higher revenues for the carriers. The most serious casualties of this decline are likely to be the demands for quality and price differentiation on the Internet.

24. Conclusions

The history of communication suggests strongly that as services become less expensive and are more widely used, the balance shifts away from the need to segment the market, and thereby to extract maximal revenues and to maximize utilization efficiency of the infrastructure. Instead, customer desire for simplicity becomes dominant.

Simplicity is likely to be much more important on the Internet than in other communication services. Customers do not care about the network, they care about their applications. Those applications are growing rapidly in number, variety, and importance, as the Internet becomes what Bill Gates has called the “digital nervous system” of more and more organizations. We will not want to worry how much to pay for a packet from site X to site Y that was generated by our request for something from site A, which then contacted site B, etc. We will be happy to pay extra for simple schemes that make our lives easy.

Flat rate is by far the simplest pricing plan, and, as predicted by Anania and Solomon [AnaniaS], it continues to dominate the data transmission market. The historical information of this paper only strengthens the arguments of Anania and Solomon and of the papers [Odlyzko7, Odlyzko8] in favor of continuing with flat rates for data transmission over core fiber optic networks. However, there are and will continue to be settings where such pricing may not be feasible. One such area is in the U.S. long distance voice telephony, where access charges are by far the largest cost component. Another such area is likely to be in wireless communication. Although the bandwidth there is growing, it is orders of magnitude lower than on fiber, and will remain orders of magnitude lower. Hence wireless bandwidth will continue to be relatively scarce (at least relative to that on fiber backbones) and technical and economic methods to ration it may continue to be required.

When usage sensitive pricing is required, customer preferences argue for only the simplest possible schemes, such as the Paris Metro Pricing proposal of [Odlyzko5]. However, it is best to avoid even schemes such as Paris Metro Pricing. There are alternatives that have a usage sensitive component, yet approximate flat rate pricing from the customer point of view. One such alternative is block pricing, which provides a user with a large allotment of time (in cases of phone calls) or bytes (for data). There are various choices that can be made with block pricing. One can charge users extra when they exceed the limits of their block purchase, thereby providing incentives for them to do what they naturally do,

namely buy more than they need. One can instead provide various controls that insure there will be no unpleasant surprises. (Either stop service when the limit is reached, or provide warnings when the limit is about to be reached,) The key point is to satisfy users' desires for simplicity, predictability, and risk avoidance.

Further along the spectrum towards true flat rate is the "expected usage pricing" of [Odlyzko8], mentioned already in the Introduction. It would be similar to the most popular Lexis/Nexis plans, with service providers offering users unlimited access for some period such as a year. The pricing would be determined by the capacity of the link and that customer's record of prior usage. Service providers would assume some growth rate in traffic, and could put into the contracts provisions for reopening them in case of unusual behavior. This type of scheme would leave scope for negotiations and for actions that improve the efficiency of the network. ("We will lower your fee by 10% if you agree to send your backups over our network at 3 in the morning, and not at 10 in the evening.") Such an approach would have several advantages for service providers. It would stimulate usage. Further, it should also reduce turnover, as a competitor attempting to attract somebody else's customer would not have the detailed knowledge of that customer, and so would face the problem of adverse selection, in which only the least desirable customers might switch.

The general conclusion is that we should strive for simplicity, even at the cost of efficiency. That is the world of communications has been evolving for the past two centuries, and that is how it is likely to evolve in the future. There will be opportunities for optimization of the network, but they will have to be pursued in ways that do not burden the end users.

The history of communication also suggests strongly that multimedia and other broadcast-like transmission will not dominate the development of the Internet. Even if there is a transitional period when most of the volume of data flows is streaming multimedia, that is not where most of the money will come from. Historically, the willingness to pay for entertainment and similar broadcast information has been low compared to that of spontaneous point-to-point communication. This suggests that the current preoccupation with "content" is misplaced, and that the future of the Internet is in enabling the myriad of connections between people, and between people and machines, connections in which most of the data flows will not be seen by any human eye, nor heard by any human ear.

Acknowledgements: I owe much to the people who have provided me with information. I am especially grateful to Jörn Altmann and his colleagues in the INDEX project for the usage data quoted in Section 9, David Cracknell for BT pricing history, David Hochfelder for quotes from Western Union archives and an unpublished paper, David Gabel for his papers, Sheldon Hochheiser for AT&T historical data and pointers to other useful sources, Richard John for his unpublished papers, Jerry Mansfield for the history of the U.S. Postal Service, Jim Reeds for telegraphic code books, and Jim Roberts for data about French systems, I also thank David Applegate, Greg Blonder, Karyen Chu, Ken Church, Whit Diffie, Ehud Gelblum, Derek Gregory, Richard Kielbowicz, Don King, Jeff Lagarias, Henry Landau, Paul Odlyzko, Alison Oswald, Philonoë, Larry Roberts, Hal Varian, Dave Walden, Roger Watt, and Barry Wellman, for comments and useful information.

Price indices, conversion rates, and data sources: All prices listed in this paper are in current dollars (i.e., not adjusted for inflation). Where non-U.S. prices are quoted, they have been converted to U.S. dollars at approximations to the exchange rate valid at the time. (For example, five dollars to one pound sterling was used for all of the 19th century.)

Widely used price indices exist for the 20th century, but the information for the 19th century is less certain. Furthermore, the standard of living has risen substantially. Therefore using cost of living indices to convert old prices into today's currency does not provide an accurate picture of the burden that prices imposed in the past. Therefore in comparing costs of communication services, it seemed better to relate them to the hours of work required to pay for them. That approach also has problems, since it is hard to find reliable wage statistics. Even when such statistics exist, they are typically for relatively small populations of urban laborers, who were not typical for a largely agrarian population. Therefore a crude procedure was used. The average wages (not counting cost of benefits) for full-time employees in the U.S. private sector in 1990 are given as \$10.01 per hour in [USDOC2]. It was assumed that the ratio of this average hourly wage to the GDP per capita stayed constant in the past. (Figures for GDP were taken from [Mitchell3], for population from [USDOC1, USDOC2].) This ignores the effects of taxes, changing family structure, and myriad other effects. However, for the crude purposes of this paper, it seemed to be adequate. The figure of a daily wage of about one dollar per day for the early 19th century that this procedure produces is in rough agreement with estimates cited on p. 144 of [Pred], for example.

Table 1.1 is based on data from [ITU]. For Japan, the data is for 1996. The revenue figures include wireless and data services. The minutes of calls per person per day figures include modem and fax

calls (which probably account for over 20% of calling minutes in the U.S.). These figures were obtained by taking the average number of minutes of calls per day (obtained from [ITU]), dividing it by the population, and multiplying it by two, since most calls involve two people simultaneously. (This treatment is inconsistent with that of other tables, such as that in Table 3.1. In that case the number of pieces of mail per person was obtained by dividing the number of pieces delivered by the population. To be consistent with Table 1.1, that figure should be doubled.) The minutes of phone calls per person statistic has shown the fastest growth in the U.S. among all the countries in this table.

Table 2.1 is based on [USDOC2] and covers data only through 1997 because for many areas, that is the most recent data available there. The growth rate is the compound annual growth rate between 1994 and 1997. The telephone industry data is from Table 927 in that work. (These figures include directory advertising revenues. They differ from those in Table 3.4 in coverage.) USPS statistics are taken from Table 946, those for advertising from Table 947, and the ones for the motion picture industry from Table 432. The airline statistics come from Table 1070 (and cover just the scheduled airline industry), those for sporting goods from Table 446. Defense spending comes from Table 574, and excludes spending on veterans benefits (\$39.3 billion in 1997). The statistics for consumer spending on “content” are derived from Table 920. The figures for TV and radio broadcasting are those for advertising in these industries in Table 947 (after excluding advertising on cable TV). Newspaper and magazine revenues in the “broadcast industries” section were obtained by adding the statistics for advertising in these media from Table 947 to the data on consumer expenditures in Table 920. The data about consumer expenditures on phone services come from the Consumer Expenditure Surveys by the U.S. Bureau of Labor Statistics, (<http://stats.bls.gov/csxhome.htm>). There is double and triple counting in this table. (Advertising spending shows up in newspaper and television revenues, for example. This spending also appears in USPS figures, since the lion’s share of the \$36.9 billion spent on direct mail advertising in 1997 went to the USPS, as well as in the phone industry figures, which received a large part of the \$11.4 spent in 1997 on yellow pages advertising.)

Table 2.2 is based on data from the National Cable Television Association, the U.S. cable TV industry association, at (<http://www.ncta.com/glance.html>), and the cellular industry association, at (<http://www.wow-com.com/wirelessurvey/>). The numbers differ from those in Table 2.1, since they come from different sources and do not cover identical services.

Expenditure figures in Table 3.1 come from [USDOC1] and from recent annual reports of the USPS. Volume statistics for the years through 1840 are derived from Table 1.2 on p. 4 of [John1]. (See also [Rich]. About half the volume in those years was newspapers, half letters.) The exception is 1810,

where this figure was obtained by doubling the number of letters delivered that year, as reported in [Miles], since there was no estimate in [John1]. For 1850, the volume estimate was obtained by taking the number of letters delivered that year that is given in [Miles] and multiplying it by an adjustment factor of 2.235 derived from the detailed breakdown of postal deliveries for the period of one year starting on July 1, 1851, that is presented on p.+107 of [Kielbowicz2]. For later years, volume figures come from [USDOC1]. Estimates for GDP and population were taken from [USDOC1] and [Mitchell3].

Table 3.2 is based on [Mitchell2]. Population figures were obtained by interpolating the values given there.

Tables 3.3 and 3.4 are derived from statistics in [USDOC1].

Table 3.4 is derived from statistics in [xxx, xxx]. The revenue figures include cell phones as well as data services.

Table 3.5 is derived from the U.S. Cellular Telecommunications Industry Association Semi-Annual Wireless Industry Survey, available at <http://www.wow-com.com/wirelessurvey/>. The subscriber numbers are for June of each year, and the revenue figures are at annualized rates.

Table 4.1 is based on data obtained from the University of Waterloo. Only the volume for the month of highest traffic in each of the three school terms each year is available.

Table 9.1 is derived from AOL quarterly reports, available at [AOL]. The quarters in Table 8.1 are calendar quarters, and thus differ in numbering from the fiscal quarters in AOL reports.

Data for Table 9.2 came from the online statistics of AFA, the association of French ISPs, which are available at <http://www.afa-france.com/html/chiffres/index.htm>.

Fig. 9.1 is based on data provided by Jörn Altmann and his colleagues in the INDEX project.

Table 12.1 is based the table on p. 348 of [Miles], and is confirmed by data in Table xxx of [Daunton], and the statistics in footnote 6 on p. 323 of [Robinson]. The figure for number of letters for 1839 does not include franks. None of the figures include newspapers, of which Rowland Hill estimated that about 30 million copies were transported in 1836, a number that grew rapidly in the 1840s. (Before the 1840 reform, newspapers were carried for free, and afterwards they enjoyed reduced rates.)

Table 12.2 is derived from [xxx].

Tables 13.1 and 13.2 are based on [xxx].

Table 13.3 is derived from the table on p. 68 of [Kieve].

Pre-1999 prices in Table 14.1 were obtained from AT&T Archives. They are the standard business rates for the shortest call that could originally be ordered, namely three minutes. Initially these were person-to-person calls, using radiotelephones. Later they became caller-dialed station-to-station calls.

Off-peak rates were introduced starting in xxx.

Table 16.1 is based on tables 1, 2, and 3 in [OECD]. The total connect time includes directory services (about one fifth of entire usage). The figures for number of terminals cover only the Minitels, and exclude PC Minitel emulation boards, which were estimated at 0.39 million in 1993 and 1.3 million in 1996 (Table 6 in [OECD]). Thus the estimate in the last column of Table 16.1 for the average time spent online per terminal, obtained from the second and third columns, is an overestimate of actual usage, especially for later years.

Table 19.1 comes from the end of 1999 online information for UUNet, available at <http://www.uu.net>. Information about pricing of other ISPs can be obtained through the online information of *Boardwatch* magazine at <http://boardwatch.internet.com>.

Table 19.2 is based on the online statistics for JANET, the British academic research network. They are available at <http://bill.ja.net>.

Table 20.1 is derived from statistics on p. 357 of [Bonjour]. The division into local and long distance is not given in that source for other types of mail, such as postcards, printed matter, or small packages.

Table 22.1 is derived from data in [FCC1].

Table 23.1 is based on listings from the January 31, 2000 weekly newsletter from RateXchange, <http://www.RateXchange.com>.

References

- [Abler] R. F. Abler, Distance, intercommunications, and geography, *Proc. Assoc. American Geographers*, 3 (1971), pp. 1–4.
- [AdamicH] L. A. Adamic and B. A. Huberman, The nature of markets in the World Wide Web, available at <http://www.parc.xerox.com/spl/groups/dynamics/topics/internetecologies.shtml>.
- [AdamsB] S. B. Adams and O. R. Butler, *Manufacturing the Future: A History of Western Electric*, Cambridge Univ. Press, 1999.
- [ActonP] J. P. Acton and R. E. Park, Response to time-of-day electricity rates by large business customers: Reconciling conflicting evidence, Report R-3477-NSF, RAND Corp., 1987.
- [Ahvenainen] J. Ahvenainen, *The Far Eastern Telegraphs: The History of Telegraphic Communication between the Far East, Europe and America before the First World War*, Ann. Acad. Scien. Fennicae #216, Helsinki, 1981.
- [AltmannRV] J. Altmann, B. Rupp, and P. Varaiya, Internet demand under different pricing schemes, in *Proc. ACM Conference on Electronic Commerce (EC-99)*, ACM, 1999, pp. 9-14.
- [Amer] S. Amer, Snubbing the Street: Could investment banks really desert Wall Street?, *Forbes ASAP*, Aug. 23, 1999, pp. 55ff. Available at <http://www.forbes.com/asap/99/0823/054.htm>.
- [AnaniaS] L. Anania and R. J. Solomon, Flat—the minimalist price, pp. 91-118 in *Internet Economics*, L. W. McKnight and J. P. Bailey, eds., MIT Press, 1997. Preliminary version in *J. Electronic Publishing*, special issue on Internet economics, <http://www.press.umich.edu/jep/>.
- [Anonym1] Anonymous, Distributing music over telephone lines, *Telephony*, vol. 18, no. 25 (Dec. 18, 1909), pp. 699-701.
- [Anonym2] Anonymous, Courting by telephone, *Telephony*, vol. 18, no. 26 (Dec. 25, 1909), p. 715.
- [AOL] AOL reports, available at www-db.aol.com/corp/news/press/.
- [Aschoff] V. Aschoff, *Geschichte der Nachrichtentechnik*, 2 volumes, Springer, 1984 and 1987.
- [ATT] American Telephone and Telegraph Company, *Telephone and Telegraph Statistics of the World*, series of annual reports.

- [Baark] E. Baark, *Lightning Wires: The Telegraph and China's Technological Modernization, 1860-1890*, Greenwood Press, 1997.
- [BaileyL] E. E. Bailey and E. B. Lindenberg, Peak load pricing principles: Past and present, pp. 9-31 in *New Dimensions in Public Utility Pricing*, H. M. Trebing, ed., MSU Public Utilities Studies, 1976.
- [Baker] S. Baker, Telefónica: Takeover escape artist? It's fending off predators with spin-offs that boost market cap, *Business Week*, April 10, 2000.
- [Baratz] M. S. Baratz, *The Economics of the Postal Service*, Public Affairs Press, 1962.
- [Barty-King] H. Barty-King, *Girdle Round the Earth: The Story of Cable and Wireless and its Predecessors, to Mark the Group's Jubilee, 1929-1979*, Heinemann, 1979.
- [Bayers] C. Bayers, Capitalist econstruction, *Wired*, March 2000, available at <http://www.wired.com/wired/archive/8.03/markets.html>).
- [Bennett] A. R. Bennett, *The Telephone Systems of the Continent of Europe*, Longmans, Green, and Co., 1895.
- [Behringer] W. Behringer, *Thurn und Taxis: die Geschichte ihrer Post und ihrer Unternehmen*, Piper, 1990.
- [Blainey] G. Blainey, *The Tyranny of Distance: How Distance Shaped Australia's History*, Macmillan, 1975. (First edition in 1968 by Sun Books.)
- [Blair] D. Blair, Impact of the Internet on core switching network, presented at ENPW '98, March 1998, Les Arcs, France. (Irish data, striking impact of lower tariffs at 6 pm.)
- [BlinderCLR] A. S. Blinder, E. R. D. Canetti, D. E. Lebow, and J. B. Rudd, *Asing About Prices: A New Approach to Understanding Price Stickiness*, Russell Sage Foundation, 1998.
- [Blonder] G. E. Blonder, Advertising Research Foundation Keynote, March 15, 1999, available at http://www.genuineideas.com/Idea_Transmogrifier/ARF%20Talk/arf-slides.htm).
- [Blondheim] M. Blondheim, *News Over the Wires: The Telegraph and the Flow of Public Information in America, 1844-1897*, Harvard Univ. Press, 1994.
- [Bonjour] E. Bonjour, ed., *Histoire des Postes Suisses, 1849-1949; Les Postes Fédérales*, 1949.

- [Braudel] F. Braudel, *Civilization and Capitalism, 15th–18th Century: Vol. 1, The Structures of Everyday Life*, Harper & Row, 1981. Original published in French in 1979.
- [Briody] D. Briody, ASPs rethink approach to billing users, *InfoWorld*, Dec. 6, 1999, p. 1.
- [Brock] G. W. Brock, *The Telecommunications Industry: The Dynamics of Market Structure*, Harvard Univ. Press, 1981.
- [Bronfman] E. Bronfman, Jr., Remarks as prepared for delivery at the Real Conference 2000, San Jose, May 26, 2000, available as a Seagram press release, (www.seagram.com/news/current-press/sc1052600b.html).
- [Brown] F. J. Brown, *The Cable and Wireless Communications of the World*, Pitman, 1927.
- [BrownD] J. S. Brown and P. Duguid, “Agents and Angels,” Chapter 2 of *The Social Life of Information*, Harvard Business School Press, 2000. Also published in *First Monday*, 5, no. 4 (April 2000), (http://firstmonday.org/issues/issue5_4/index.html).
- [BrownS] S. J. Brown and D. S. Sibley, *The Theory of Public Utility Pricing*, Cambridge Univ. Press, 1986.
- [BrynjolfssonS] E. Brynjolfsson and M. Smith, Frictionless commerce? A comparison of Internet and conventional retailers, 1999 report. Available at (<http://ecommerce.mit.edu/erik/>).
- [Cairncross1] F. Cairncross, A survey of telecommunications: The death of distance, *The Economist*, Sept. 30, 1995. Available at (<http://www.economist.com/>).
- [Cairncross2] F. Cairncross, *The Death of Distance: How the Communications is Revolution will Change our Lives*, Harvard Business School Press, 1997.
- [Castells] M. Castells, *The Rise of the Network Society*, Blackwell, 1996.
- [Chandler] A. D. Chandler, Jr., *The Visible Hand: The Managerial Revolution in American Business*, Harvard Univ. Press, 1977.
- [Chu] K. Chu, User reactions to flat-rate options under time charges with differentiated quality of access: Preliminary results from INDEX, available at (<http://www.marengoresearch.com/isqe/>).

- [ChuangS] J. C.-I. Chuang and M. S. Sirbu, Optimal bundling strategy for digital information goods: Network delivery of articles and subscriptions, *Information Economics and Policy*, 11 (1999), pp. 147–176.
- [Coase1] R. H. Coase, The nature of the firm, *Economica* (N.S.) 4 (1937), pp. 386-405.
- [Coase2] R. H. Coase, Rowland Hill and the Penny Post, *Economica* 6 (Nov. 1939), pp. 423-435.
- [Coase3] R. H. Coase, The economics of uniform pricing systems, *The Manchester School*, 15 (May 1947), pp. 139-156.
- [Coase4] R. H. Coase, “Talking about tomorrow,” interview in the special Jan. 1, 2000 edition of *The Wall Street Journal*, “The Next 1000 Years,” available at <http://interactive.wsj.com/millennium>).
- [CoatesF] V. T. Coates and B. Finn, *A Retrospective Technology Assessment: Submarine Telegraphy, The Transatlantic Cable of 1866*, San Francisco Press, 1979.
- [Coe] L. Coe, *The Telegraph.; A History of Morse’s Invention and its Predecessors in the United States*, McFarland & Co., 1993.
- [CoffmanO1] K. G. Coffman and A. M. Odlyzko, The size and growth rate of the Internet. *First Monday*, Oct. 1998, <http://firstmonday.org>. Also available at <http://www.research.att.com/~amo>.
- [CoffmanO2] K. G. Coffman and A. M. Odlyzko, Internet growth: Is there a “Moore’s Law” for data traffic?, manuscript in preparation.
- [CosgroveL] J. G. Cosgrove and P. B. Linhart, Customer choices under local measured telephone service, *Public Utilities Fortnightly*, August 30, 1979, 27-31.
- [Cowell] A. Cowell, Correspondent’s Report: Service slips, fares baffle on British trains, *New York Times*, May 28, 2000 (Travel Section). Available at <http://www.nytimes.com/00/05/28/travel/ta000528.html>.
- [Cracknell1] D. Cracknell, Growing the telecommunications market—lessons from the past, presented at ITS Regional Meeting, Crete, Sept. 1994.
- [Cracknell2] D. Cracknell, The impact of structural changes in telephony prices, pp. 97–109 in *Beyond Competition: The Future of Telecommunications*, D. Lamberton, ed., Elsevier, 1995.

- [Cracknell3] D. Cracknell, Telephone usage and demographic trends: An analysis of changes in UK call demand 1986-1997, presented at ITS Conference, Stockholm, June 1998.
- [CracknellK] D. Cracknell and M. Knott, The measurement of price elasticities—the BT experience, *Intern. J. Forecasting*, vol. 11 (1995), pp. 321–329.
- [Cressy] D. Cressy, *Coming Over: Migration and Communication between England and New England in the Seventeenth Century*, Cambridge Univ. Press, 1987.
- [CrewK1] M. A. Crew and P. R. Kleindorfer, Rowland Hill’s contribution as an economist, pp. 1-11, with Comments by M. J. Daunton, pp. 13-16, in [CrewK2].
- [CrewK2] M. A. Crew and P. R. Kleindorfer, eds., *Competition and Innovation in Postal Services: Proceedings of a Conference in Honor of Rowland Hill*, Kluwer, 1991.
- [CrewK3] M. A. Crew and P. R. Kleindorfer, eds., *Managing Change in the Postal and Delivery Industries*, Kluwer, 1997.
- [CrewK4] M. A. Crew and P. R. Kleindorfer, eds., *Emerging Competition in Postal and Delivery Industries*, Kluwer, 1999.
- [CTIA] CTIA (Cellular Telecommunications Industry Association), April 11, 2000 press release, “CTIA Reports 1999 Survey Results,” available at http://www.wow-com.com/news/ctiapress/body.cfm?record_id=857).
- [Daunton] M. J. Daunton, *Royal Mail: The Post Office Since 1840*, The Athlone Press, 1985.
- [DeneckereM] R. J. Deneckere and R. P. McAfee, Damaged goods, *J. Economics and Management Strategy*, vol. 5, no. 2 (1966), pp. 149-174.
- [deSolaP1] I. de Sola Pool, ed., *The Social Impact of the Telephone*, MIT Press, 1977.
- [deSolaP2] I. de Sola Pool, *Forecasting the Telephone: A Retrospective Technology Assessment*, Ablex, 1983.
- [deSolaPITH] I. de Sola Pool, H. Inose, N. Takasaki, and R. Hurwitz, *Communications Flows: A Census in the United States and Japan*, North-Holland, 1984.
- [Dibner] B. Dibner, *The Atlantic Cable*, Burndy Library, 1959.

- [Dordick] H. S. Dordick, *The Information Society: A Retrospective View*, Sage, 1993.
- [Douglas] S. J. Douglas, *Inventing American Broadcasting*, Johns Hopkins Univ. Press, 1987
- [DuBoff1] R. B. DuBoff, Business demand and the development of the telegraph in the United States, 1844-1860, *Business History Review* 54 (1980), 459-479.
- [DuBoff2] R. B. DuBoff, The telegraph and the structure of markets in the United States, 1845-1890, *Research in Economic History* 8 (1983), 253-277.
- [Durlacher] Durlacher Research Ltd., Quarterly Internet Report Q4/99, executive summary available at <http://www.durlacher.com/research/res-iqreportsq499execsummary.asp>).
- [EdellV] R. Edell and P. Varaiya, Providing Internet access: What we learn from INDEX, *IEEE Network*, vol. 13, no. 5 (1999), pp. 18-25. Also available at http://www.path.berkeley.edu/~varaiya/papers_ps.dir/networkpaper.pdf).
- [Ettema] J. S. Ettema, Interactive electronic text in the United States: Can videotext ever go home again?, pp. 105-123 in *Media Use in the Information Age: Emerging Patterns of Adoption and Consumer Use*, J. L. Salvaggio and J. Bryant, eds., Lawrence Erlbaum Associates, 1989.
- [EU] Commission of the European Communities, Commission recommendation on leased lines interconnection pricing in a liberalised telecommunications market, Nov. 24, 1999, Brussels, document C(1999)3863, available at <http://www.ispo.cec.be/infosoc/telecompolicy/en/Main-en.htm>).
- [FCC1] U.S. Federal Communications Commission, Trends in Telephone Service, Sept. 1999. Available from <http://www.fcc.gov/ccb/stats>).
- [FCC2] U.S. Federal Communications Commission, Trends in the U.S. International Telecommunications Industry, Sept. 1999. Available from <http://www.fcc.gov/ccb/stats>).
- [Ferguson] N. Ferguson, *The House of Rothschild : The World's Banker, 1849-1999*, Viking, 1999.
- [FergusonH] P. Ferguson and G. Huston, *Quality of Service: Delivering QoS on the Internet and in Corporate Networks*, Wiley, 1998.
- [Field1] A. J. Field, The magnetic telegraph, price and quantity data, and the new management of capital, *J. Economic History* 52 (1992), 401-413.
- [Field2] H. M. Field, *The Story of the Atlantic Cable*, Scribner's, 1893.

- [Fischer1] C. S. Fischer, Touch someone: The telephone industry discovers sociability, 1876-1940, *Technology and Culture* 29 (1988), pp. 32-61.
- [Fischer2] C. S. Fischer, *America Calling: A Social History of the Telephone to 1940*, Univ. California Press, 1992.
- [FishburnO] P. C. Fishburn and A. M. Odlyzko, Dynamic behavior of differential pricing and Quality of Service options for the Internet, pp. 128–139 in *Proc. First Intern. Conf. on Information and Computation Economies (ICE-98)*, ACM Press, 1998. Available at <http://www.research.att.com/~amo>.
- [FishburnOS] P. C. Fishburn, A. M. Odlyzko, and R. C. Siders, Fixed fee versus unit pricing for information goods: competition, equilibria, and price wars, *First Monday*, vol. 2, no. 7 (July 1997), <http://www.firstmonday.org/>. Also to appear in *Internet Publishing and Beyond: The Economics of Digital Information and Intellectual Property*, B. Kahin and H. Varian, eds., MIT Press, 2000. Available at <http://www.research.att.com/ amo>.
- [Flichy] P. Flichy, *Dynamics of Modern Communication: The Shaping and Impact of New Communication Technologies*, Sage, 1995. Original French version published in 1991.
- [Friedlander1] A. Friedlander, *Emerging Infrastructure: The Growth of Railroads*, Corp. for National Research Initiatives, 1995.
- [Friedlander2] A. Friedlander, *Natural Monopoly and Universal Service: Telephones and Telegraphs in the U. S. Communications Infrastructure, 1837-1940*, Corp. for National Research Initiatives, 1995.
- [Friedlander3] A. Friedlander, *Power and Light: Electricity in the U.S. Energy Infrastructure, 1870-1940*, Corp. for National Research Initiatives, 1996.
- [Friedlander4] A. Friedlander, “*In God We Trust*”; *All Others Pay Cash: Banking as an American Infrastructure, 1800-1935*, Corp. for National Research Initiatives, 1996.
- [Friedman] W. F. Friedman, *The History of Codes and Code Language, the International Telegraph Regulations Pertaining Thereto, and the Bearing of this History on the Cortina Report*, U. S. Government Printing Office, 1928. Reprinted as *History of the Use of Codes* by Aegean Park Press.

- [FullertonK] D. Fullerton and T. Kinnaman, Household responses to pricing garbage by the bag, *Am. Econ. Rev.*, vol. 86, no. 4 (Sept. 1996), pp. 971-984.
- [Gabel1] D. Gabel, *The Evolution of a Market: The Emergence of Regulation in the Telephone Industry of Wisconsin, 1893-1917*, Ph.D. Dissertation, University of Wisconsin–Madison, 1987.
- [Gabel2] D. Gabel, Private telecommunications networks: An historical perspective, pp. 35-49 in *Private Networks, Public Objectives*, E. Noam and A. Ni'shu'illeabha'in, eds., Elsevier, 1996,
- [GabelW] D. Gabel and D. F. Weiman, Historical perspectives on competition and interconnection between local exchange companies, pp. 75-106 in *Opening Networks to Competition: The Regulation and Pricing of Access*, D. Gabel and D. F. Weiman, eds., Kluwer, 1998.
- [GaleK] W. A. Gale and R. Koenker, Pricing interactive computer services, *Computer Journal*, vol. 27, no. 1 (1984), pp. 8-17.
- [Garfinkel] S. Garfinkel, *Database Nation: The Death of Privacy in the 21st Century*, O'Reilly & Associates, 2000.
- [GarfinkelL1] L. Garfinkel and P. B. Linhart, The transition to local measured telephone service, *Public Utilities Fortnightly*, August 16, 1979, 17-21.
- [GarfinkelL2] L. Garfinkel and P. B. Linhart, The revenue analysis of local measured telephone service, *Public Utilities Fortnightly*, October 9, 1980, 15-21.
- [GazzaleMKM] R. S. Gazzale and J. K. MacKie-Mason, System design, user cost and electronic usage of journals, available at (<http://www.si.umich.edu/PEAK-2000/program.htm>).
- [Gibbs] W. W. Gibbs, Transportation's perennial problems, *Scientific American*, 277, no. 4 (Oct. 1997), pp. 54-57. Available at (<http://www.sciam.com/1097issue/1097gibbs.html>).
- [Goldhaber] M. H. Goldhaber, The attention economy and the Net, *First Monday*, vol. 2, no. 4 (April 1997), (<http://firstmonday.org/>).
- [Greenberg] B. S. Greenberg, Teletext in the United Kingdom: Patterns, attitudes, and behaviors of users, pp. 87-101 in *Media Use in the Information Age: Emerging Patterns of Adoption and Consumer Use*, J. L. Salvaggio and J. Bryant, eds., Lawrence Erlbaum Associates, 1989.

- [Gregory] D. Gregory, The friction of distance? Information circulation and the mails in early nineteenth-century England, *J. Historical Geography*, 13 (1987), 130-154.
- [GuptaSW] A. Gupta, D. O. Stahl, and A. B. Whinston, The Internet: A future tragedy of the commons?, available at (<http://cism.bus.utexas.edu/res/wp.html>).
- [Haldi] J. Haldi (in association with J. F. Johnston, Jr.), *Postal Monopoly: An Assessment of the Private Express Statutes*, American Enterprise Institute, 1974.
- [Hansell] A. Hansell, Now AOL everywhere: Internet's giant prepares to leap off the desktop, *New York Times*, July 4, 1999.
- [Harlow] A. F. Harlow, *Old Wires and New Waves: The History of the Telegraph, Telephone, and Wireless*, D. Appleton-Century, 1936.
- [HausmanN1] W. J. Hausman and J. L. Neufeld, Time-of-day pricing in the U.S. electric power industry at the turn of the century, *Rand J. Economics* 15 (1984), 116-126.
- [HausmanN2] W. J. Hausman and J. L. Neufeld, Engineers and economists: Historical perspectives on the pricing of electricity, *Technology and Culture* 30 (1989), pp. 83-104.
- [Headrick] D. R. Headrick, *The Invisible Weapon: Telecommunications and International Politics, 1851-1945*, Oxford Univ. Press, 1991.
- [Hill] P. Hill, *The Post Office of Fifty Years Ago: Containing Reprint of Sir Rowland Hill's Famous Pamphlet, Dated 22nd February, 1837, Proposing Penny Postage: With Facsimile of the Original Sketch for the Postage Stamp, and Other Documents*, Cassell & Co., 1887.
- [Hochfelder] D. Hochfelder, A comparison of the postal telegraph movement in Great Britain and the United States, 1866-1920, *Enterprise and Society*, (2000), to appear.
- [HolzmannP] G. Holzmann and B. Pehrson, *The Early History of Data Networks*, IEEE, 1994.
- [Hughes] T. P. Hughes, *Networks of Power: Electrification in Western Society, 1880-1930*, Johns Hopkins Univ. Press, 1983.
- [Hunter] K. Hunter, PEAK and Elsevier Science, available at (<http://www.si.umich.edu/PEAK-2000/program.htm>).

- [Inglis] K. S. Inglis. The imperial connection: Telegraphic communication between England and Australia, 1872-1902, pp. 21–38 in *Australia and Britain: Studies in a Changing Relationship*, A. F. Madden and W. H. Morris-Jones, eds., Frank Cass, 1980.
- [Irvin] D. R. Irvin, Modeling the cost of data communication for multi-node computer networks operating in the United States, *IBM J. Res. Develop.* vol. 37 (1993), pp. 537-546.
- [Isard] W. Isard, *Methods of Regional Analysis*, MIT Press, 1960.
- [Israel] P. B. Israel, *From Machine Shop to Industrial Laboratory: Telegraphy and the Changing Context of American Invention, 1830-1920*, Johns Hopkins Univ. Press, 1992.
- [ITU] World Telecommunications Indicators Database (5th ed.), ITU, 1998. Available at <http://www.itu.int/publications/bookstore.html>.
- [JacksonCW] D. C. Jackson, W. H. Crumb, and G. W. Wilder, *Report on the Telephone Situation in the City of Chicago; In Respect to Service, Rates, Regulation of Rates, etc.; submitted to The Committee on Gas, Oil and Electric Light of the City Council of the City of Chicago*, Gunthorp-Warren Printing Co., 1907.
- [Janelle] D. G. Janelle, Spatial reorganization: A model and concept, *Ann. Assoc. American Geographers*, 59 (1969), pp. 348–364.
- [Jepsen] T. C. Jepsen, *Ma Kiley: The Life of a Railroad Telegrapher*, Texas Western Press, 1997.
- [John1] R. R. John, *Spreading the News: The American Postal System from Franklin to Morse*, Harvard Univ. Press, 1995.
- [John2] R. R. John, The politics of innovation, *Daedalus*, vol. 127, no. 4 (Fall 1998), pp. 187-214.
- [John3] R. R. John, Recasting the information infrastructure for the Industrial Age, pp. 55-105 in *A Nation Transformed by Information: How Information has Shaped the United States from Colonial Times to the Present*, A. D. Chandler, Jr., and J. W. Cortada, eds., Oxford Univ. Press, 2000. To appear.
- [JohnstonGS] R. J. Johnston, D. Gregory, and D. M. Smith, *The Dictionary of Human Geography*, 3rd ed., Blackwell, 1994.

- [Jordan] B. W. Jordan, The Allegheny City water works, 1840-1907, *Western Pennsylvania Historical Magazine*, vol. 70 (Jan. 1987), pp. 29-52.
- [Judson1] C. H. Judson, Unprofitable traffic—What shall be done with it?, *Telephony*, vol. 18, no. 24 (Dec. 11, 1909), pp. 644-647.
- [Judson2] K. B. Judson, ed., *Selected Articles on Government Ownership of Telegraph and Telephone*, Wilson, 1914.
- [Kahn] D. A. Kahn, *The Codebreakers; The Comprehensive History of Secret Communication from Ancient Times to the Internet*, rev. ed., Scribner, 1996.
- [Kanfer] A. Kanfer, It's A Thin World: The Association Between Email Use and Patterns of Communication and Relationships, March 10, 1999 report, available at http://www.ncsa.uiuc.edu/edu/trg/info_society.html.
- [KatzA] J. E. Katz and P. Aspden, A nation of strangers?, *Comm. ACM* 40 (1997), 81-86.
- [Kelly] F. P. Kelly, Notes on effective bandwidth, pp. 141-168 in *Stochastic Networks: Theory and Applications*, F. P. Kelly, S. Zachare, and I. Ziedins, eds., Royal Statistical Soc. Lecture Notes Series, vol. 4, Oxford Univ. Press, 1996.
- [Kielbowicz1] R. B. Kielbowicz, *Origins of the Second-Class Mail Category and the Business of Polycymaking, 1863-1879*, Assoc. for Education in Journalism and Mass Communication, Journalism Monographs, no. 96, 1986.
- [Kielbowicz2] R. B. Kielbowicz, *News in the Mail: The press, Post Office, and Public Information, 1700-1860s*, Greenwood Press, 1989.
- [Kieve] J. L. Kieve, *The Electric Telegraph: A Social and Economic History*, David and Charles, 1973.
- [Klopfenstein] B. Klopfenstein, Problems and potential of forecasting the adoption of new media, pp. 21–41 in *Media Use in the Information Age: Emerging Patterns of Adoption and Consumer Use*, J. L. Salvaggio and J. Bryant, eds., Lawrence Erlbaum Associates, 1989.
- [KowalskiW] J. Kowalski and B. Warfield, Modelling traffic demand between nodes in a telecommunications network, Proc. ATNAC 95, available at <http://www.elec.uow.edu.au/conferences/advpapers.html#RTFToC5>.

- [Kraepelien] H. Y. Kraepelien, Local telephone pricing: The configuration theory, U.S. Dept. Commerce, National Tech. Info. Service, report PB-280 194, March 1978.
- [Krause] J. Krause, Global Crossing plans its media play, *The Standard*, March 27, 2000. Available at <http://www.thestandard.com/article/display/0,1151,13209,00.html>.
- [KrautLPKM] R. Kraut, V. Lundmark, M. Patterson, S. Kiesler, T. Mukhopadhyay, and W. Scherlis, Internet paradox: A social technology that reduces social involvement and psychological well-being?, *American Psychologist* 53 (1998), 1017-1031. Available at <http://www.apa.org/journals/amp/amp5391017.html>.
- [Lay] M. G. Lay, *Ways of the World: A History of the World's Roads and of the Vehicles That Used Them*, Rutgers Univ. Press, 1992.
- [Lesk] M. Lesk, How much information is there in the world?, 1997 unpublished paper, available at <http://www.lesk.com/mlesk/diglib.html>.
- [Licklider] J. C. R. Licklider, *Libraries of the Future*, MIT Press, 1965.
- [LickliderV] J. C. R. Licklider and A. Veza, Applications of information technology, *IEEE Proc.* 66 (1978), 1330-1346.
- [Lipartito] K. Lipartito, *The Bell System and Regional Business; The Telephone in the South, 1877-1920*, The Johns Hopkins Univ. Press, 1989.
- [Lockhart] J. G. Lockhart, *Memoirs of the Life of Sir Walter Scott*, vol. 4, Baudry's, 1837-1838.
- [Lu] C. Lu, *The Race for Bandwidth: Understanding Data Transmission*, Microsoft Press, 1998.
- [Lubrano] A. Lubrano, *The Telegraph: How Technology Innovation Caused Social Change*, Garland, 1997.
- [MacKieMV1] J. K. MacKie-Mason and H. R. Varian, Pricing congestible network resources, *IEEE J. Selected Areas Comm.*, 13 (1995), 1141-1149. Available at <http://www.sims.berkeley.edu/~hal/people/hal/papers.html>.
- [MacKieMV2] J. K. MacKie-Mason and H. R. Varian, Economic FAQs about the Internet, pp. 27-62 in [McKnightB].

- [MajumdarRD] R. C. Majumdar, H. C. Raychaudhuri, and K. Datta, *An Advanced History of India*, MacMillan, 1946.
- [MargolisR] M. Margolis and D. Resnick, Third Voice: Vox populi vox Dei, *First Monday*, October 1999, (http://firstmonday.org/issues/issue4_10/margolis/index.html).
- [Marvin] C. Marvin, *When Old Technologies Were New: Thinking About Electric Communication in the Late Nineteenth Century*, Oxford Univ. Press. 1990.
- [McKnightB] L. W. McKnight and J. P. Bailey, eds., *Internet Economics*, MIT Press, 1997. Preliminary version of many papers available in J. Electronic Publishing, special issue on Internet economics, (<http://www.press.umich.edu/jep/>).
- [Mendelson] H. Mendelson, Pricing computer services: Queueing effects, *Comm. ACM*, 28 (1985), pp. 312-321.
- [Miles] P. Miles, History of the Post Office, *The Bankers' Magazine, and Statistical Register*, vol. 7 (new series, sometimes listed as vol. 12 in the original series), no. 6 (Nov. 1857), 337-365.
- [Mitchell1] B. M. Mitchell, Pricing policies in selected European telephone systems, pp. 437-475 in *Proc. Sixth Annual Telecommunications Policy Research Conference*, D. C. Heath, 1979.
- [Mitchell2] B. M. Mitchell, *International Historical Statistics, Europe, 1750-1993*, 4-th ed., Macmillan, 1998.
- [Mitchell3] B. M. Mitchell, *International Historical Statistics, The Americas, 1750-1993*, 4-th ed., Macmillan, 1998.
- [MitchellV] B. M. Mitchell and I. Vogelsang, *Telecommunications Pricing: Theory and Practice*, Cambridge Univ. Press, 1991.
- [Mueller] M. L. Mueller, Jr., *Universal Service: Competition, Interconnection, and Monopoly in the Making of the American Telephone System*, MIT Press and AEI Press, 1997.
- [MuellerS] M. L. Mueller and J. R. Schement, Universal service from the bottom up: A study of telephone penetration in Camden, New Jersey, *The Information Society* 12, no. 3 (1996), 273-292.

- [NixG] J. Nix and D. Gabel, AT&T's strategic response to competition: Why not preempt entry?, *J. Economic History* 53, no. 2 (1993), 377-387.
- [Noll1] A. M. Noll, *Introduction to Telephones and Telephone Traffic*, 2nd ed., Artech House, 1991.
- [Noll2] A. M. Noll, Does data traffic exceed voice traffic?, *Comm. ACM*, June 1999, pp. 121-124.
- [Odlyzko1] A. M. Odlyzko, Tragic loss or good riddance? The impending demise of traditional scholarly journals, *Intern. J. Human-Computer Studies* (formerly *Intern. J. Man-Machine Studies*) 42 (1995), pp. 71-122. Also in the electronic *J. Univ. Comp. Sci.*, pilot issue, 1994, <http://hyperg.iicm.tu-graz.ac.at>. Available at <http://www.research.att.com/~amo>.
- [Odlyzko2] A. M. Odlyzko, The bumpy road of electronic commerce, in *WebNet 96 - World Conf. Web Soc. Proc.*, H. Maurer, ed., AACE, 1996, pp. 378-389. Available at <http://www.research.att.com/~amo>.
- [Odlyzko3] A. M. Odlyzko, The slow evolution of electronic publishing, pp. 4-18 in *Electronic Publishing '97: New Models and Opportunities*, A. J. Meadows and F. Rowland, eds., ICC Press, 1997. Available at <http://www.research.att.com/~amo>.
- [Odlyzko4] A. M. Odlyzko, Smart and stupid networks: Why the Internet is like Microsoft, *ACM netWorke* 2(5) (Dec. 1998), 38-46. Available at <http://www.research.att.com/~amo>.
- [Odlyzko5] A. M. Odlyzko, Paris Metro Pricing for the Internet, in *Proc. ACM Conference on Electronic Commerce (EC-99)*, ACM, 1999, pp. 140-147. Based on a 1997 unpublished manuscript, A modest proposal for preventing Internet congestion. Both available at <http://www.research.att.com/~amo>.
- [Odlyzko6] A. M. Odlyzko, Data networks are lightly utilized, and will stay that way. Available at <http://www.research.att.com/~amo>.
- [Odlyzko7] A. M. Odlyzko, The economics of the Internet: Utility, utilization, pricing, and Quality of Service. Available at <http://www.research.att.com/~amo>.
- [Odlyzko8] A. M. Odlyzko, The Internet and other networks: Utilization rates and their implications. Presented at the 1998 Telecommunications Policy Research Conference. To appear in *Information Economics & Policy*. Available at <http://www.research.att.com/~amo>.

- [Odlyzko9] A. M. Odlyzko, The visible problems of the invisible computer: A skeptical look at information appliances, *First Monday*, vol. 4, no. 9 (Sept. 1999), (<http://www.firstmonday.org/>). Also available at (<http://www.research.att.com/~amo>).
- [OECD] OECD, France's experience with the Minitel: Lessons for electronic commerce over the Internet, report DSTI/ICCP/IE(97)10/FINAL, available at (<http://www.oecd.org/dsti/sti/it/infosoc/prod/minitel.htm>).
- [Oslin] G. P. Oslin, *The Story of Communications*, Mercer Univ. Press, 1992.
- [Palmeri] C. Palmeri, Bully Bell, *Forbes*, April 22, 1996, pp. 92, 94.
- [Panzar] J. C. Panzar, The Pareto domination of usage-insensitive pricing, pp. 425-436 in *Proc. Sixth Annual Telecommunications Policy Research Conference*, D. C. Heath, 1979.
- [ParkM] R. E. Park and B. M. Mitchell, Optimal peak-load pricing for local telephone calls, Report R-3401-1-RC, RAND Corp., 1987.
- [Parsons] F. Parsons, *The Telegraph Monopoly*, C. F. Taylor, 1899.
- [Philonoe] Philonoe, At large and at small: Mail, *The American Scholar*, vol. 69, no. 1, Winter 2000, pp. 7-11.
- [Platt] H. L. Platt, The cost of energy: Technological change, rate structures, and public policy in Chicago, 1880-1920, *Urban Studies* 26 (1989), pp. 32-44.
- [Pred] A. R. Pred, *Urban Growth and the Circulation of Information: The United States System of Cities, 1790-1840*, Harvard Univ. Press, 1973.
- [Read] D. Read, *The Power of News: The History of Reuters, 1849-1989*, Oxford Univ. Press, 1992.
- [Reeds] J. A. Reeds, Commercial telegraphic code books, unpublished manuscript and description of database, available at (<http://www.research.att.com/~reeds>).
- [ReichLS] P. Reichl, S. Leinen, and B. Stiller, A practical review of pricing and cost recovery for Internet services, to appear in Proc. 2nd Internet Economics Workshop Berlin (IEW'99), Berlin, Germany, May 28-29, 1999. Available at (<http://www.tik.ee.ethz.ch/~cati/>).
- [Reid] J. D. Reid, *The Telegraph in America. Its Founders, Promoters, and Noted Men*, Derby Brothers, 1879.

- [Rich] W. E. Rich, *The History of the United States Post Office to the Year 1829*, Harvard Univ. Press, 1924.
- [Robinson] H. Robinson, *The British Post Office: A History*, Princeton Univ. Press, 1948.
- [SabatinoA] R. Sabatino and J. M. de Arce, Implementation of the Pan-European academic research network: TEN-155, paper presented at the TERENA-NORDUnet Networking Conference 1999. Available at <http://www.terena.nl/tnnc/proceedings.html>.
- [SchaferV] A. Schafer and D. Victor, The past and future of global mobility, *Scientific American*, vol. 277, no. 4 (Oct. 1997), pp. 58-61. Available at <http://www.sciam.com/1097issue/1097schafer.html>.
- [Scheele] C. H. Scheele, *A Short History of the Mail Service*, Smithsonian Institution Press, 1970.
- [Schlender] B. Schlender, Sony plays to win, *Fortune*, May 1, 2000, pp. 142+. Available at <http://www.fortune.com/fortune/2000/05/01/mak.html>.
- [Schwarzlose] R. A. Schwarzlose, *The Nation's Newsbrokers, vol. 2: The Rush to Institution, from 1865 to 1920*, Northwestern Univ. Press, 1988.
- [ShapiroV] C. Shapiro and H. Varian, *Information Rules: A Strategic Guide to the Network Economy*, Harvard Business School Press, 1998.
- [Smith] A. D. Smith, *The Development of Rates of Postage; An Historical and Analytical Study*, G. Allen & Unwin, 1918.
- [Smulyan] S. Smulyan, *Selling Radio: The Commercialization of American Broadcasting, 1920-1934*, Smithsonian Institution Press, 1994.
- [Sorkin] A. L. Sorkin, *The Economics of the Postal Business: Alternatives and Reform*, Lexington Books, 1980.
- [StArnaud] B. St Arnaud, The future of the Internet is NOT multimedia, *Network World*, Nov. 1997. Available at <http://tweetie.canarie.ca/~bstarn/publications.html>.
- [StArnaudCFM] B. St Arnaud, J. Coulter, J. Fitchett, and S. Mokbel, Architectural and engineering issues for building an optical Internet. Short version in *Proc. Soc. Optical Engineering*, (1998). Full version available at <http://www.canet2.net>.

- [Standage] T. Standage, *The Victorian Internet: The Remarkable Story of the Telegraph and the Nineteenth Century's On-line Pioneers*, Walker, 1998.
- [Stehman] J. W. Stehman, *The Financial History of the American Telephone and Telegraph Company*, Houghton Mifflin, 1925.
- [Swisher] K. Swisher, *Aol.Com: How Steve Case Beat Bill Gates, Nailed the Netheads, and Made Millions in the War for the Web*, Times Books, 1998.
- [Szabo] N. Szabo, Micropayments and mental transaction costs, to appear in proceedings of the 2nd Berlin Internet Economics Workshop. Available at (<http://www.best.com/~szabo/micropay.ps>).
- [Taggart] S. Taggart, Telstra: The prices fight, *Wired News*, (<http://www.wired.com/news/politics/0,1283,32961,00.html>).
- [TarrFG] J. A. Tarr, T. Finholt, and D. Goodman, The city and the telegraph: Urban communications in the pre-telephone era, *J. Urban History* 14 (1987), 38–80.
- [Taylor] G. R. Taylor, *The Transportation Revolution, 1815-1860*, Rinehart, 1951.
- [Temin] P. Temin with L. Galambos, *The Fall of the Bell System: A Study in Prices and Politics*, Cambridge Univ. Press, 1987.
- [Thompson] R. L. Thompson, *Wiring a Continent: The History of the Telegraph Industry in the United States, 1832-1866*, Princeton Univ. Press, 1947.
- [USDOC1] U.S. Department of Commerce, Bureau of the Census, *Historical Statistics of the United States: Colonial Times to 1970*, 1975.
- [USDOC2] U.S. Department of Commerce, Bureau of the Census, *Statistical Abstract of the United States 1999*, 1999. Available online at (<http://www.census.gov/prod/www/statistical-abstract-us.html>).
- [Varian1] H. R. Varian, A model of sales, *Am. Economic Review*, 70 (1980), pp. 651-659. Erratum on p. 517 of vol. 71 (1981).
- [Varian2] H. R. Varian, Price discrimination, pp. 597-654 in *Handbook of Industrial Organization*, vol. I, R. Schmalensee and R. D. Willing, eds., Elsevier, 1989.
- [Varian3] H. R. Varian, *Microeconomic Analysis*, 3rd ed., Norton, 1992.

- [Varian4] H. R. Varian, Differential pricing and efficiency, *First Monday*, vol. 1, no. 2 (Aug. 1996), [⟨http://firstmonday.org/⟩](http://firstmonday.org/).
- [Varian5] H. R. Varian, Pricing information goods, available at [⟨http://www.sims.berkeley.edu/~hal/people/hal/papers.html⟩](http://www.sims.berkeley.edu/~hal/people/hal/papers.html).
- [Webb] H. L. Webb, *The Development of the Telephone in Europe*, Electrical Press, 1910.
- [Weber] R. E. Weber, Seward's other folly: The fight over America's first encrypted cable, *Cryptologia* 19 (1995), 321-348.
- [WelfensY] P. J. J. Welfens and G. Yarrow, eds., *Telecommunications and Energy in Systemic Transformation: International Dynamics, Deregulation and Adjustment in Network Industries*, Springer, 1997.
- [Wellman] B. Wellman, Changing connectivity: A future history of Y2.03K, *Sociological Research Online*, Sociological Research Online, vol. 4, no. 2, Feb. 2000. Available at [⟨http://www.chass.utoronto.ca/~wellman/links/index.html⟩](http://www.chass.utoronto.ca/~wellman/links/index.html).
- [Wells] D. A. Wells, *The Relation of the Government to the Telegraph; or, A Review of the Two Propositions Now Pending Before Congress for Changing the Telegraphic Service of the Country*, 1873.
- [Wenders] J. T. Wenders, Two views of applied welfare analysis: The case of local telephone service pricing, *Southern Economic J.* 57 (1990), 340-348.
- [Wilson1] A. G. Wilson, *Urban and Regional Models in Geography and Planning*, Wiley, 1974.
- [Wilson2] G. Wilson, *The Old Telegraphs*, Phillimore, 1976.
- [Winston] B. Winston, *Media Technology and Society: A History: From the Telegraph to the Internet*, Routledge, 1998.
- [Wolak] F. A. Wolak, Changes in the household-level demand for postal delivery services from 1986 to 1994. Preliminary version available at [⟨http://www.stanford.edu/~wolak/⟩](http://www.stanford.edu/~wolak/).
- [Zajac] E. E. Zajac, *Political Economy of Fairness*, MIT Press, 1995.
- [Zipf] G. K. Zipf, Some determinants of the circulation of information, *Amer. J. Psychology* 59 (1946), 401-421.