Critical Business Decisions for Internet Services

Joseph P. Bailey

Ph.D. Candidate, Technology, Management and Policy
Massachusetts Institute of Technology
77 Massachusetts Avenue, Bldg. E40-237
Cambridge, MA 02139-4307 U.S.A.
+1-617-253-2373 (tel)
+1-617-253-7326 (fax)
bailey@rpcp.mit.edu

Lee W. McKnight

Lecturer, Technology and Policy Program
Principal Investigator, Internet Telephony Consortium
Massachusetts Institute of Technology
77 Massachusetts Avenue, Bldg. E40-235
Cambridge, MA 02139-4307 U.S.A.
+1-617-253-0995
mcknight@rpcp.mit.edu

Husham S. Sharifi

Technology and Policy Program
Research Assistant, Internet Telephony Consortium
Massachusetts Institute of Technology
77 Massachusetts Avenue, Bldg. E40-218
Cambridge, MA 02139-4307 U.S.A.
+1-617-253-6828
sharifi@rpcp.mit.edu

Prepared for presentation at the 1998 Information Resources Management Association International Conference,

Global Information Technology Management Track

May 17-20, 1998

Radisson Hotel, Boston, Massachusetts

INTRODUCTION

Critical business and user decisions greatly affect the development of the Internet infrastructure. Because the Internet marketplace is competitive and decentralized, an Internet Service Provider or other business' Internet interconnection agreements determines user performance, cost structure, and ultimately profitability. As the Internet evolves from best effort services to integrated services, the interconnection agreements become even more important.

The Internet is characterized by decentralization and growth. Decentralization allows for heterogeneity of users and applications ranging from electronic mail to the World Wide Web. Decentralization has also led to heterogeneous network designs and scalable interconnection agreements. The Internet continues to grow as measured by the number of users and hosts, the amount and types of traffic, the advertising volumes and transactions, and the size of the Internet industry. With continued growth, some expect the Internet to become even more fragmented, although others see evidence of a trend towards consolidation.[1] Save for a few centralized functions such as allocation of Internet Protocol (IP) addresses, Domain Name Service (DNS), and distribution of routing tables, the Internet has consistently followed a decentralized administration philosophy.[2] Plans for decentralization of some of these functions are being discussed.[3]

This article examines the ability of Internet interconnection models[4] to incorporate technological innovations such as integrated services.[5] The point of interconnection is an illuminating focus for Internet analysis. Interconnecting networks must cooperate to exchange information to help the Internet function. Interconnection points must also relay the bits and bytes of an e-mail message or a Web session across different networks. Assessment by Internet Service Providers (ISPs) of a range of technology, policy, management, and economic factors determines how and where interconnections occur. Integrated and differentiated services on the Internet represent an important development since they make possible new applications, such as real-time video or voice. Integrated services also fundamentally change the economics of Internet interconnection agreements, and more broadly require new business models, and likely a new industrial structure and regulatory framework to enable growth.[6]

This article presents four models of interconnection agreements for Internet Service Providers and describes interconnection point pricing policies. To anchor these theories within the reality of the current interconnection marketplace, the article includes empirical analysis of the evolving Internet interconnection business environment. The article continues by detailing the emergence of integrated and differentiated services. Finally, drawing upon incomplete contracts theory, an integrated analysis of these elements is provided.[7] By integrating the technology of the Internet (interconnection systems and integrated services) and the economics of interconnection agreements (pricing policies and incomplete contract theory), we determine which interconnection models will be sustainable with alternative integrated services pricing policies. This analysis suggests that usage-sensitive pricing policies, which have been very rare in the Internet community to date, will become more pervasive as adoption of integrated services grows. Along with this development, new accounting, billing, network management, test and measurement, and signaling methods must be implemented. We conclude that only by adopting new pricing policies will the Internet become scalable to new services, new applications, and new users.

This future Internet may rely on the market for Bearer Service provision as described by Kavassalis, Lee, and Bailey (1997). Unlike the current Internet Protocol (version 4), the bearer service will emerge to support interactive multimedia services along with today's "best effort" Internet traffic. While a deeper discussion of the Bearer Service is beyond the scope of this article, the marketplace for Bearer Service provision may be susceptible to greater price competition.[8] The analysis in this article regarding interconnection of networks providing integrated services is extensible to the Bearer Service market.

INTERCONNECTION AGREEMENTS

Interconnection of networks that exchange Internet Protocol (IP) traffic is the glue that holds together the Internet as a "network of networks." However, interconnection occurs between numerous types of networks at many different locations. Many different companies and user groups can be affected by the terms of interconnection agreements. Although sending an e-mail message or a Web transmission between two computers can span numerous networks (and therefore will take advantage of the existence of many interconnection agreements), the details of the interconnection agreement do not have to be known by those people who use it. Rather, it is up to the connecting parties to decide how they wish to interconnect with other networks.

In this article, we describe four basic models of Internet Service Provider interconnection agreements: peer-to-peer, hierarchical, intermediary, and cooperative agreements. Each of these models exists in the Internet. It is very likely that all of these types of interconnection agreements will exist in the future. An interesting feature of the models, and of the Internet, is that different models for interconnection can coexist and interoperate without one model dominating. This permits networks and firms with different technologies and customers to select the interconnection agreement that suits them best. The four interconnection models are described in detail below.

Peer-to-Peer

The peer-to-peer interconnection model consists of two Internet networks owned by different firms of approximately the same size, experience, technology, and customer base that interconnect via a two-party contract governing their agreement. An example of this model is the interconnection of two Internet Service Providers that both have a national reach and are of similar size. Even though these providers are competing in the same market for the same customers, they must provide interconnection between their networks so that their customers will realize the benefits of positive network externalities.[9]

The network externalities are symmetric in the peer-to-peer agreement since both networks have approximately the same customer base. Although the users are the ones who actually derive the benefit from the network externality, the value of a network (and therefore the price that can be set for being connected to the network) is proportional to the network externalities that it has. Consider an ISP that is deciding whether to enter a peer-to-peer agreement with either a company of equal size or with one that is half its size. If the ISP connects to the one of equal size, it will double the number of users to which it can directly connect through this interconnection point. If the ISP connects to the one half its size, then it is only increasing the number of possible direct connections by 50 percent. Both connections may connect the ISP with the Internet "cloud,"[10] but connection with the network that is half its size will make more of the connections *indirect*. Clearly, direct connections are more desirable than indirect connections since traffic must traverse fewer links and must pass through fewer intermediaries. Therefore, having indirect connections would make the indirect traffic susceptible to the reliability of the intermediary network. However, it may still make sense for the ISP to connect to the smaller network in a hierarchical bilateral agreement, which we discuss below.

The elements that make two firms peers as opposed to placing them into more of a customer-provider relationship are important: size, experience, technology, and customer base. The customer base and its implications are apparent in the discussion of network externalities above. The size of a network is very important since one network may have a national reach while another may have only a regional reach. All else being equal, an interconnection agreement with a network which has a larger reach is better than one with a small reach since larger networks can help transport packets a farther distance and reduce the number of networks a packet must traverse (i.e., it decreases the "indirectness" of connections to other Internet locations).[11] Experience is very important since the parties entering the interconnection agreement must be able to trust each other to successfully transport the data that is exchanged between them. If the interconnecting parties have asymmetric experience, the party with greater experience (or knowledge) may act as a mentor or teacher to the less experienced party. A similar argument can be made for technology. If there is a large difference in the level of technological development, the capacity of the network with more developed technology to benefit from the interconnection is less.[12] This effect is magnified by the introduction of new services such as integrated services.

Hierarchical

Similar to the peer-to-peer agreement, the hierarchical agreement is governed by a two-party contract, but it interconnects firms of a discernible difference in size, experience, technology, and customer base. We distinguish this bilateral agreement from the peer-to-peer case since the economics and technology of the interconnection are very different. Both differences usually lead to a customer-provider relationship rather than a peer-to-peer relationship. An example of this interconnection model is an Internet Access Provider or a corporate network connecting to an ISP.

The technology leader, usually the firm with the larger network, has less to benefit from the interconnection agreement than the technology follower. The leader may also have more experience or network links with greater capacity. The network externality benefits are greater for customers of the smaller network provider. The experience of the technology leader can be of great benefit to the firm with the smaller network, but the experience and knowledge benefits do not flow in the opposite direction. For these reasons, we find that the firm with a smaller network takes on the role of the customer and pays a larger amount of the costs of interconnection.

The hierarchical model of interconnection is the most pervasive in today's Internet. Customers connect to Internet Access Providers. Internet Access Providers and corporate networks connect to Internet Service

Providers. All of these interconnection agreements follow the hierarchical model since they aggregate users and interconnect them with networks which have superior technology and larger network externalities.

Intermediary

The intermediary model is followed when an interconnection point consists of more than two networks exchanging packets and the administration of the interconnection is operated by a firm that does not operate a network. Examples of this include the Commercial Internet eXchange (CIX), the Network Access Points (NAPs) established by the National Science Foundation, and MAE-E (pronounced “may east”). The roles of the intermediary are to route traffic between the interconnected networks and to serve as a trusted party to facilitate communication and promote nondiscrimination. Because of these roles, an intermediary often acts as a common carrier, offering consistent prices to all customers and not refusing interconnection by any party. The objective of the intermediary is to cover the operating expenses of the interconnecting points and profit from the endeavor.

The network externalities for the intermediary are characterized by positive feedback. If the number of people connecting to the intermediary’s interconnection point is zero, the first network to connect receives zero benefit. As the number of networks that connect to the interconnection point increases, the network externalities also increase. It is difficult, therefore, for an intermediary to establish an interconnection point, but once it attains a critical mass of firms, it can provide a very real benefit to new networks. A wise strategy for these intermediaries would be to secure a critical mass before establishing an interconnection point.

The intermediary must establish trust, which results partially from a technological edge. Since the intermediary does not usually compete directly with the networks it interconnects, it does not benefit from information it obtains about its customers (the interconnecting networks). Furthermore, the intermediary is likely to keep this information about its customer confidential since it benefits from keeping this information private — this is the intermediary’s competitive advantage. Sharing information in confidence establishes the intermediary’s trustworthiness. The intermediary’s ability to impart some kind of experience or knowledge to their customers also builds trust. An intermediary that knows less than its customers is unlikely to provide adequate service.

Cooperative Agreement

Similar to the intermediary agreement, the cooperative agreement has more than two parties sharing an interconnection point; however, with a cooperative agreement, the operation of the interconnection point is run by a committee of the interconnecting firms. This interconnection model was the sole example of interconnection when the Internet comprised government-supported networks. The Federal Internet eXchanges (FIXen) were created to interconnect government agency networks that had a shared purpose and incentive to promote research and education. While it is unclear that this model works in the commercial sector, it is worth mentioning, since the FIXen still exist, remain successful, and similar, cooperative interconnections operate in other nations. Unlike the intermediary, the cooperative agreement is run by committee and does not need to make a profit—only cost-sharing is necessary.

The cooperative agreement is more desirable than many bilateral agreements since there are fewer coordination costs and greater economies of scale. This is especially true because there is an incentive alignment between the parties involved that makes them willing to cooperate and have very incomplete contracts (discussed further below), and use their trust in each other to facilitate coordination. Also, multiple bilateral agreements are not only more expensive to coordinate, but may cost more in terms of hardware and leased lines than a single interconnection point of n networks. This gives the cooperative agreement an economies-of-scale benefit that is the same as the economies of scale exhibited by the intermediary.

Since information is not a competitive advantage for the cooperative firms, experience and information are shared among all parties involved. Instead of information being absorbed by the administrator, all parties can benefit from committee participation and shared learning.

PRICING POLICIES

Pricing for Internet service is one of the most critical issues to emerge from the commercialization of the Internet following the NSFNET. Two of the first articles to suggest pricing based upon actual usage initiated a debate over the cost recovery and business of providing Internet service.[13] Economists and technologists continue to debate, and disagree, on how and whether to implement new pricing policies.[14] There are still more research and potential business opportunities to explore to further our understanding of pricing and Internet economics.

In this section we describe three pricing strategies: flat-rate, capacity-based, and usage-sensitive pricing. These pricing strategies coexist today within and across market segments, and our analysis that follows indicates that they will continue to coexist in the future Internet infrastructure.

Flat-Rate Pricing

The pricing approach that most people think prevails on the Internet today is flat-rate pricing. With flat-rate pricing, a user does not pay for each transmission of data but only has to pay for the initial cost of the connection, typically through a fixed, monthly subscription charge. The flat rate is set independent of the speed or configuration of the connection.[15] Because configuration and speed of connection are very important in a heterogeneous Internet, we find that flat-rate pricing is not as common as pricing based on capacity. In fact, when carefully examined, *flat-rate* is a misleading term for what could be more accurately thought of as a capacity-based, fixed-rate pricing policy. That is, there is a capacity limit on the flat-rate user—one may use up to 144 kbps, but no more in the case of an ISDN line. ISDN is often priced on a simple usage basis, which makes it unattractive to those with an alternative.

One of the benefits of flat-rate pricing is that it is easy to set prices and bill for services, and consequently, for users to budget for their expected bills. No accounting is necessary to track usage. Also, no price discrimination has to be determined for pricing based on the speed of connection. Flat-rate pricing is easy to administer and easy for customers to understand. Furthermore, it encourages usage since users do not have to pay any additional fee to upgrade the speed (bandwidth) of their connection or reduce their usage because of billing, up to the aforementioned limits.

Capacity-Based Pricing

Capacity-based pricing relates pricing to usage by setting a price based on the bandwidth or speed of the connection. This policy is based on the *expected* use of a circuit since no accounting is done on the link. This is accomplished by charging for the configuration (i.e., bandwidth) of the connection but not the actual bits sent and/or received. Capacity-based pricing differentiates classes of users (price discrimination), so network providers can charge more where there is more benefit.

Capacity-based pricing is the currently-prevailing pricing policy on the Internet. While it requires a more complicated billing policy than does flat-rate pricing, it helps the network provider increase its revenue. As users or other networks connect with bigger (i.e., greater bandwidth) connections, they are charged more because they can increase congestion on the shared portion of the network. The larger revenues from higher-capacity links help cover the costs of infrastructure improvements.

A weak point of capacity-based pricing is that it does not track the customer's actual use. The customer may be inflicting a higher congestion cost on the network when he or she sends and/or receives traffic during the middle of the day instead of using the network during the off-peak evening hours. Simple capacity-based pricing does not provide an incentive feedback mechanism to change the behavior of the customer by encouraging use of the network at night. Furthermore, not all customers that have the same capacity connection use the network in the same way. Some users prefer the low latency that a larger capacity link offers, but do not use it often (many periods of zero traffic on the link). Other users fill their capacity and have less sporadic use. Both types of customer would pay the same price in a capacity-based pricing model. It would take a usage-sensitive pricing model to differentiate these users.

Usage-Sensitive Pricing

Unlike capacity-based pricing, usage-sensitive pricing policies may also include charges for *actual* usage rather than only for expected usage. This distinction is important because usage-sensitive pricing requires accounting as well as billing, whereas capacity-based pricing only requires billing for cost recovery.

How the accounting is implemented for usage-sensitive pricing may determine pricing, and thus the incentives for users and network providers as well. For example, if accounting is done at the connection level (i.e., TCP layer), then billing for network traffic will also be done at that level.[16] Accounting for packets at the IP layer will lead to billing for packets, not bits. Accounting and billing will then be reported back to the users who may in turn change their behavior to save money.

Usage-sensitive pricing gives the network provider more flexibility when designing its price model. Pricing by time of day or priority of traffic will help the network provider reduce congestion on its network by shaping the behavior of users. This is, in essence, a form of load management. The current difficulty with this kind of pricing policy is that users are not used to a variable price for their Internet service (note that capacity-based pricing would have the same cost per billing cycle regardless of use). It is difficult for people at this early point in the development and diffusion of Internet services to accept usage-sensitive pricing when capacity-based pricing is easier (and requires less accounting overhead from the network provider).

However, usage-sensitive pricing is possible and does exist. For example, in New Zealand an expensive link to the United States is supported by charging users based upon their actual use of the network.[17] Users pay for the bits they send and receive. It is a nonlinear pricing scheme with discounts for high-volume users which reflects the fact that there is an overhead for billing users that does not change much with their actual use (the cost is just averaged over more volume). For best effort service, we believe the New Zealand case is a unique example, at the time of this writing, of a successful usage-sensitive pricing policy. In places where there is competition based on pricing policies, it is less clear what usage-sensitive pricing policies will be acceptable to users. What is clear, however, is that such long-distance interconnection will prevail for some time to come. The consequent results that are sketched in the section that follows are unexpected.

THE GLOBAL INTERNET INTERCONNECTION INDUSTRY STRUCTURE

While there are only a few theoretical models for interconnection, there are a wide array of different points to which a network may connect. Table 1 contains a listing of these points around the world. Note that these connection points are all governed by an intermediary model. Several networks also connect with each other bilaterally, although outside the U.S. such phenomena often result in interconnection decisions which defy the previous, and apparently dying, importance of distance. Examples of such geographically illogical interconnection are offered in this section.

United States (NAPs)	Europe (NAPs)	Rest of World (NAPs)	Metropolitan Area Ethernets (MAEs)
Vermont Internet Exchange (VIX) Boston MXP (Boston's official exchange) Global NAPs (another Boston-based exchange) Philadelphia Internet Exchange Baltimore NAP (BNAP) Telehouse America Internet Packet Exchange (New Jersey based) ATL-NAP (Atlanta based) Atlanta Internet eXchange Commercial Internet Exchange Chicago: Bellcore and Ameritech San Francisco Bay Area: Bellcore and Pacific Bell New York: Sprint Washington DC: MFS SIX (Seattle Internet eXchange) EWIX (Eastern Washington/Idaho Internet Exchange) LAP (Los Angeles) PAIX (Digital's Palo Alto exchange) OIX (Oregon Internet Exchange) SNNAP (Puget Sound Regional Interconnection) NIX (Northwest Internet eXchange) St. Louis	LINX -- London Internet Exchange FICIX -- Finnish Internet Exchange AMS -IX – Amsterdam Internet Exchange DH-IX -- Den Haag Regional Internet Exchange CIXP – CERN Exchange for Europe (Geneva) SIX (Swiss Exchange) DIX - Danish Internet Exchange deCIX (Deutsche CIX) dGIX (a Nordic Exchange) INEX (Ireland Neutral Exchange) DINX (Irish Data Internet Neutral Exchange) Manchester NAP BIX (Budapest Internet eXchange) Slovak Internet eXchange) Neutral Internet Exchange (Czech Republic) VIX (Vienna Internet eXchange) SFINX (Paris) MIXITA (Milan) NAP/Roma (Rome) ESPANIX (Madrid) PIX (Portuguese Internet eXchange point) Iceland BNIX (Belgium) M9-IX (Moscow)	LIX – Latin Internet Exchange, Dominican Republic CA/NAP (Toronto) MIX (Montreal Internet Exchange) QIX (Quebec Internet Exchange) Vancouver, BCTel HKIX (Hong Kong) JPIX (Japan) NSPIXP-2 (Japan) STIX (Singapore) WAIX (Western Australia Internet Exchange) AUIX (Australian Internet Exchange) NZIX (New Zealand) MM-MAP (Philippines) InteRED (Panama) Embratel (Brazil) LNCC (Brazil) GT-ER (Brazil) IIX (Israel) South Africa	MAE-Houston AMAP (Austin) MAE -East -- Washington DC MAE- West -- San Jose, CA MAE Chicago MAE Dallas MAE Houston MAE New York MAE Paris, France MAE Frankfurt, Germany

MXP (Detroit)	SPB-IX (St. Petersburg)
IndyX (Indianapolis Data Exchange)	AIX (Athens Internet Exchange)
Ohio Exchange	
NNAP (Nashville)	
Mountain Area eXchange (Denver)	
Zion’s Hill (Utah)	
The ARCH (St. Louis)	
Pennsuken	
Houston	
TTN (The Tucson NAP, non-NSF)	
TTI (The Tucson Interconnect)	
Phoenix Exchange (MPIX)	
New Mexico Network Access Point	
FIX – Federal Internet Exchange	

Table 1. NAPs and MAEs around the world.[18]

There is a wide variety of pricing levels for these interconnection points as well. A NAP offering only regional interconnection, such as the Utah REP (Regional Exchange Point), prices 100 Mbps service far below the same service offered by a NAP that provides national backbone connectivity, such as that in Los Angeles. Other NAPs exist purely as cooperative efforts and consequently offer lower bandwidth service for free. Table 2 offers a sampling of representative connection points.

PacBell	**Ameritech**	**New Jersey MAE**	**LAP (Los Angeles)**	**Utah REP**	**SIX (Seattle)**
$5500, DS3	$4000, DS3	$1250, T1	$4500, 100 Mbps	$25, 10 Mbps	Free, 10 Mbps
$6956, OC3	$5900, OC3			$60, 100 Mbps	

Table 2. Pricing of selected interconnection points: cost per month.

Interconnection between networks in the U.S. occurs through a wide variety of schemes, mostly depending on the size of the networks involved, as measured by the number of routes with a certain capacity, and the traffic load. All the previously mentioned types of interconnection are employed, although some to a much greater extent than others (as was mentioned for the hierarchical arrangement). In fact, other than essentially theoretical debates over the types of interconnection which lead to social optimality,[19] the actual interconnection of networks in the U.S. is working well enough that few people approach the topic contentiously.

Interconnection agreements outside the U.S. remain controversial. One cannot, for example, analyze the European interconnection situation in the same way as one would in the United States—namely, by focusing on NAPs. There are three primary reasons for this, two of which are directly related to the NAPs and one of which arises from the character of the European backbones.

The first issue is that European NAPs are smaller than U.S. NAPs. There are some exceptions to this observation. For example, in England and in Scandinavia large NAPs are either present or emerging. But even these are large only within the European context, meaning that they provide connectivity of 155 Mbps at the most. This is a mere 10% of the capacity of large U.S. NAPs. With such limitations, network backbones that handle the amount of traffic which is typical for today's Internet do not have realistic continental options for NAP interconnection.

But even if they had, perhaps they would still not connect. For instance, to date Deutsche Telekom, the largest ISP in Europe, will not connect to the NAP in Frankfurt. They have made this decision despite the peering of 25 other German ISPs at the same point. Deutsche Telekom instead chooses to peer with Deutsches Forschungsnetz, a German academic network, thus leaving the commercial option effectively commercially dead.[20] The implications are even worse for German data. If there does not exist some common peering point between Deutsches Forschungsnetz and either the Frankfurt NAP or one of the 25 mentioned ISPs, the only way packets from Deutsche Telekom could travel to these other ISPs is by taking circuitous routes—a system that happens with IP routers anyway, but not to such extremes. It is even possible, and likely, that for a packet to travel between two regions within Germany, the packet would first travel outside of Germany and then back.

A controversial question is where outside the local region would the packet travel. After all, in any general scenario, a packet travels only through the interconnections of the network from which its data session has been initiated. The controversial and real-life answer is that in most cases the packet passes through one of the U.S. backbone networks.

This occurs precisely because of the discrepancy in size and reliability between the U.S. Internet and the networks of the rest of the world. Europe, which is the best of the bunch outside of the U.S., has decidedly “thin” backbone networks by U.S. standards. A recent Yankee Group report entitled "The European Internet: Enter the Telco" has found that of Europe’s five major backbone providers, only two have pervasive 34 Mbps routes.[21] As is common in Europe, the other providers have routes as thin as 2 Mbps in certain parts of the network. These numbers offer a sharp contrast with the U.S., where 622 Mbps backbones are typical and where short term plans for expansion have goals of doubling and trebling of bandwidth.

The situation is even more extreme in Asia and elsewhere. In some cases, not only are the regional networks unable to carry high bandwidth traffic, but the connections to bigger networks, such as those in the U.S., are desperately inadequate. India offers an illustrative example, with just 10 Mbps of connectivity to the U.S. for the entire country to share.[22]

The results are that in order to move traffic within their region (e.g., intra-European traffic, intra-Asian traffic), many ISPs outside the U.S. link to a U.S. backbone. Thus, the bandwidth bottleneck experienced by Internet users often is occurring on the trans-oceanic link. The above scenarios lead to situations where large networks could peer according to intermediary models in their own region but instead choose to peer in other regions (or, more specifically, across the world) according to a peer-to-peer scheme in some cases and a hierarchical scheme in most cases.

In short, when thinking of interconnection issues outside of the U.S., one should first consider the cost of interconnecting with the U.S. A CEO of one of the major European backbones claims that even the largest players look at the issue within this framework. In this contextual light, a sample of pricing for a hierarchical connection from Europe to the U.S. is presented in Table 3.

Bandwidth of Link	**One year lease (per month charge)**	**Two year lease (per month charge)**	**Three year lease (per month charge)**
1024 kbps	$15,435	$14,663	$13,340
1544 kbps	$19,624	$18,643	$16,960
2048 kbps	$23,152	$22,457	$21,800

Table 3. Half Link Prices for International Private Lease Line between Europe and the U.S.

Table 3 describes the price of connection between Geneva, Switzerland and the MCI backbone in the U.S. There are several extremely important points to note. First, the minimum lease term is one year. Second, the prices shown are, as indicated, monthly charges. Third, the service is provided by MCI but still must pass through Swiss Telecom's local network. Fourth, this implies that the above numbers do not include the local lines payments which Swiss Telecom would necessarily receive. Finally, the costs shown actually represent only half of the price that the network in Geneva would need to pay MCI. They are only half-link costs, meaning that they buy the right to lease only one of the two links in the connection. The numbers are listed in this fashion by MCI presumably as a result of traditional interconnection agreements between PSTN networks, in which a bilateral interconnection would stipulate that each network pays half the link cost. But the network leasing a data line from Geneva to the U.S. pays for the link on its end and on the North American end, despite the two-way flow of data traffic. The issue has caused considerable international protests, with network owners outside the U.S. claiming that this type of interconnection constitutes an implicit subsidy of the U.S. backbones. And yet these same networks—which benefit much more from the robust U.S. backbones than do the U.S. backbones from them—still choose to interconnect, thus ensuring that the debate will continue for some time to come.

The consequences of this infrastructure setup are noteworthy, but they do not explicitly address the magnitude of money flows. The central issue is the margin between the cost of deploying international private lines to the U.S. backbones and the price the U.S. backbones charge to lease them. In brief, the margin is huge. Owning a so-called Minimum Investment Unit (MIU) on trans-Pacific fiber from the U.S. to Australia, for example, costs about $12,638 per month for network management, billing, and general overhead (i.e. variable costs). The largest cost is the fixed cost of purchasing and installing the capacity. That is the cost that AT&T, just as an example, would incur. The price they charge to lease that same 2 Mbps is $98,000 per month. The profit margin is nearly 90% and is sustainable since the large fixed cost of entry is very high.

Whether these prices are a result of market clearing or of market imperfection is debatable. Lines to the U.S. are leased in greater numbers with each passing day. The important consideration is whether this trend will continue in the absence or in the presence of regional development, likely accelerated by the January 1st (1998) deregulation of European Union local access markets. Or perhaps the trend will simply shift away from U.S. interconnection. The operators of the networks must consider these exogenous events when deploying, even though they have highly limited information with which to forecast. This is the essence of incomplete contract theory—and it can be effectively used by operators to understand how, within the above constraints, they may gain an edge by deploying such technologies as Integrated Services.

INTEGRATED SERVICES

One of the greatest benefits of the Internet is its ability to incorporate dynamically new protocols and technologies.[23] Attention has focused on technical, business, and user requirements for the Internet to provide integrated and differentiated services.[24] Since not all applications have the same bandwidth requirements, the integrated services model allows for different applications to call on different protocols to deliver different qualities of service. The current quality of service, best effort, is seen as basic integrated service, while guaranteed service (suitable for a lossless, constant bit rate transmission) is the most demanding. While today the Internet can only deliver very low-quality video or voice, a future Internet may be able to offer a guaranteed quality of service that is consistent with the public switched telephone network. The capacity of integrated services to offer better quality seems likely with development of the new Internet Protocol version 6 (IPv6).

This section of the article concentrates on the reservation process for providing integrated and differentiated services, such as guaranteed service. A reservation is the set-up procedure for allocating bandwidth from the source to destination. It conveys a priori information about the data stream to tell the network to set aside capacity. Every interconnection point through which the traffic passes must have the reservation supported in order to provide true guaranteed quality of service. At the time of writing, protocols for reservations such as RTP and RSVP are currently being developed and tested.

Reservations affect the interconnection analysis because not all traffic is handled equally. The ability of these protocols to reserve bandwidth on networks other than their own is necessary to ensure consistent quality, but may be inconsistent with the business models of Internet Service Providers. For example, consider a hierarchical agreement in which the larger network reserves bandwidth on the smaller network. It could be possible for the larger network to reserve *all* of the bandwidth on the smaller network for integrated services.

ANALYSIS

By determining which pricing policies are consistent with different interconnection agreements in adopting integrated and differentiated services, we will now combine the models for interconnection agreements, pricing policies, and the introduction of integrated services. Our methodology will be analysis through incomplete contract theory.[25]

Incomplete contracts are a result of a firm's bounded rationality. It is impossible for people or firms to know everything that will happen in the future; they are bounded by their rational thinking. However, they are aware of this when they enter into a contract. Thus by definition, the contract will not take into consideration all future events. The contract cannot be complete since events exogenous to the contract are unknown and firms participating in the contract may act beyond the limits of the contract. Analysis stemming from incomplete contract theory looks at the exogenous changes affecting the contract and determines where a firm can act opportunistically.

This is a useful methodology for this analysis since the Internet and consequently, the interconnections between networks are very dynamic. The development of new applications like the World Wide Web change the use and behavior of the network but have little effect on the actual protocols that affect interconnection. The introduction of integrated services, however, has a large impact on interconnection since it makes it possible to impose a large opportunity cost on an interconnected party by reserving bandwidth on its network. Furthermore, network providers can find new ways to act opportunistically in an interconnection agreement since they often compete with the firm with which they are interconnecting.

Table 4 summarizes the analysis already presented in this article and presents the findings of the analysis below. The first row of Table 4 lists examples of the different interconnection models. The second row indicates the nature of the network externality benefits for each model (this analysis is also described earlier in this article). The third row indicates which pricing policies are possible for best effort service for the various interconnection models.[26] Furthermore, the accounting overhead associated with usage-sensitive pricing is too great for a cooperative agreement, since the incentive for all interconnected parties is very much in alignment. Therefore, the flat-rate, intermediary and the usage-sensitive cooperative model are not sustainable, scalable interconnection agreements, as the third row of Table 4 shows.

	Peer-to-Peer	*Hierarchical*	*Intermediary*	*Cooperative Agreement*
Examples	ISP-ISP	IAP-ISP	MAE-E	FIX
Network Externalities	mutually beneficial	asymmetric - more beneficial to the smaller firm	positive feedback	positive feedback

Best Effort Pricing Policies	capacity-based, usage-sensitive, and flat-rate	capacity-based, usage-sensitive, and flat-rate	usage-sensitive, capacity-based	capacity-based, flat-rate
Reservation Pricing Policies	usage-sensitive	usage-sensitive	usage-sensitive	usage-sensitive
Costs	equal split	smaller firm pays more	all firms pay total cost + profit	all firms pay total cost
Agreement	contract	contract	common carrier	committee
Accounting	possible	possible	probable	not likely

Table 4. Comparison of Interconnection Models.

We now turn our attention to the last four rows of Table 4 (reservation pricing policies, costs, agreement, and accounting), which summarize the analysis in this article.

As the fourth row of Table 4 indicates, flat-rate or capacity-based pricing schemes are consistent with any of the three interconnection models when integrated services are introduced. The ability of any one user to reserve bandwidth on any other network means imposing a very high congestion cost (in some cases, taking away the entire capacity of the network) without incurring any cost (under a flat-rate or capacity-based pricing scheme). Only a usage-sensitive pricing scheme can prevent a tragedy of the commons.[27] This is especially true in cases where the interconnecting parties are competing, but it may also be true in cooperative agreements. For example, while the National Science Foundation and NASA might agree on a cooperative interconnection, they may have two users who do not get along (rarely does the U.S. government agree on anything!). One user could establish a reservation without incurring a cost and take away the bandwidth from the other user.

Another issue that arises is the capability of usage-sensitive pricing to be based on the current congestion level of a network. While this is only one method for setting a price, it has been suggested before.[28] It offers incentives to create artificial congestion on a network so that the network can raise its price and increase its revenues. This is less likely in the cooperative case, but it may be likely in all the other cases since the artificial congestion is promoted by a party that stands to make more money. The intermediary may be less likely to create artificial congestion since it does not own a network, and it is trusted more than peers trust each other (since they are competing in the same markets).

The table's last three rows (costs, agreement, and accounting) detail differences between the parties in the interconnection agreement. Most of the points highlighted in this table are included earlier in this article. Here, we wish to highlight the last row—accounting. Arguably, accounting is the most probable outcome in the intermediary case for best effort service, since intermediaries are likely to be the first to adapt usage-sensitive pricing. Therefore, this model will incur the lowest costs in a transition to reservation policies.

CONCLUSION

Internet interconnection agreements are as heterogeneous as the users of and the traffic on the networks connected by the Internet. Changes in interconnection agreements and pricing strategies are evident as new services such as integrated and differentiated services, including a variety of interactive multimedia communication services, are introduced. However, no single model of interconnection must dominate on the Internet. What is essential is for interconnection agreements to become scalable to accommodate delivery of integrated services. When users can establish reservations or signal preferences, and potentially, tremendously increase the congestion costs of other users, a scalable interconnection framework to support end-to-end delivery of integrated services will be required. Business decisions on the parameters of usage-sensitive pricing strategies may be the critical element for ensuring that interconnection frameworks for integrated services are scalable for new applications.

ACKNOWLEDGMENTS

The authors thank David Clark, John Wroclawski, William Lehr, and Scott Marcus for their contributions. This work has been supported in part by the Internet Telephony Consortium (http://itel.mit.edu) and NASA (fellowship no. NGT-51407), the National Science Foundation (grant no. NCR-9307548). The views expressed, as well as any errors of fact or omission are those of the authors, and not their sponsoring institutions.

REFERENCES

Bailey, J.P., “The Emergence of Cybermediaries: Intermediation in Electronic Markets,” MIT Ph.D. Dissertation, forthcoming.

Bailey, J.P., “The Economics of Internet Interconnection Agreement,” *Internet Economics*, L.W. McKnight and J.P. Bailey, eds. MIT Press, Cambridge, Mass., 1997.

Bailey, J.P. and L.W. McKnight, "Scalable Internet Interconnection Agreements and Internet Integrated Services," *Coordinating the Internet*, B. Kahin and J. Keller, eds. MIT Press, Cambridge, Mass., 1997.

Bohn, R., H.-W. Braun, et al., “Mitigating the Coming Internet Crunch: Multiple Service Levels via Precedence,” *Applied Network Research* technical report, 1994.

Brownlee, N., “New Zealand's Experiences with Network Traffic Charging,” *Internet Economics*, L.W. McKnight and J.P. Bailey, eds. MIT Press, Cambridge, Mass., 1997.

Clark, D.D., “The Design Philosophy of the DARPA Internet Protocols,” *Computer Communication Review*, Vol. 18, No. 4, 1988, pp. 106-114.

Coase, R.H., “The Nature of the Firm,” *Economica*, 1937, Vol. 44, pp. 386-405.

Edell, R.J., N. McKeown, et al., “Billing Users and Pricing for TCP,” University of California at Berkeley, unpublished paper, 1994.

Evagora, A., “World Wide Weight,” *tele.com*, September 1997, available at *http://www.teledotcom.com/0997/features/tdc0997globe.html*.

Farrell, J. and G. Saloner, “Competition, Compatibility and Standards: The Economics of Horses, Penguins and Lemmings,” *Product Standardization and Competitive Strategy*, H.L. Gabel, ed. North-Holland, New York, 1987.

Gillett, S. and M. Kapor, “The Self-Governing Internet: Coordination by Design,” *Coordinating the Internet*, B. Kahin and J. Keller, eds. MIT Press, Cambridge, Mass., 1997, pp. 3-38.

Hardin, G., “The Tragedy of the Commons,” *Science*, Vol. 162, No. 13, 1968, pp. 1243-1248.

International Telecommunication Union, “Generic Top Level Domains Memorandum of Understanding (gTLD-MoU), February 28, 1997, available at *http://www.gtld-mou.org/gTLD-MoU.html*.

Katz, M.L. and C. Shapiro, “Network Externalities, Competition, and Compatibility” *American Economic Review*, Vol. 75, No. 3, 1985, pp. 424-440.

Kavassalis, P., T. Lee, and J. Bailey, “Sustaining an Independent Bearer Service Market,” Internet Telephony Consortium, Massachusetts Institute of Technology, 1997.

Kelly, Tim, H. Shawn Sharifi, and Ben Petrazinni, *Challenges to the Network: Telecoms and the Internet*, ITU Press, Geneva, Switzerland, 1997.

MacKie-Mason, J.K. and H.R. Varian, “Pricing the Internet,” *Public Access to the Internet*, B. Kahin. and J. Keller, eds. MIT Press, Cambridge, Mass., 1995, pp. 269-314.

Maloff, J., *1994-1995. Internet Access Providers Marketplace Analysis*, The Maloff Company, 1995.

McKnight, L.W. and J.P. Bailey, *Internet Economics*, MIT Press, Cambridge, Mass., 1997.

McKnight, L.W. and J.P. Bailey, “Global Internet Economics,” *Brazilian Electronic Journal of Economics*, December 1997, available at *http://www.beje.decon.ufpe.br/mcknight.htm*.

McKnight, L.W. and J.P. Bailey, “Internet Economics: When Constituencies Collide in Cyberspace,” *IEEE Internet Computing*, Vol. 1, No. 6, November-December 1997, pp. 30-37.

Rekhter, Y., P. Resnick, et al., “Pricing for Internet Addresses and Route Assignments,” Telecommunications Policy Research Conference, Solomons Island, MD, October 6, 1996.

Shenker, S., “Service Models and Pricing Policies for Integrated Services Internet,” Xerox PARC, unpublished paper, 1993.

Williamson, O., *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, New York, 1975.

NOTES

* An earlier version of this article appeared as Bailey, J. and L. McKnight. 1997. “Scalable Internet Interconnection Agreements and Integrated Services,” in Kahin, B. and J. Keller, eds. Coordinating the Internet. Cambridge, MA: MIT Press, 309-324.

---

[1] Maloff, *1994-1995. Internet Access Providers Marketplace Analysis*, (1995).

[2] Clark, “The Design Philosophy of the DARPA Internet Protocols” (1988).

[3] See, for example, the Memorandum of Understanding on the Generic Top Level Domain Name Space of the Internet Domain Name System (gTLD-MoU), International Telecommunication Union (February 28, 1997), available at < http://www.gtld-mou.org/gTLD-MoU.html>; and Rekhter, Resnick, et al., “Pricing for Internet Addresses and Route Assignments” (October 6, 1996).

[4] Bailey, “The Economics of Internet Interconnection Agreements” (1997).

[5] Shenker, “Service Models and Pricing Policies for Integrated Services Internet” (1993).

[6] In McKnight, Bailey, eds. (1997), many of these issues were addressed. See the chapters by Gong and Srinagesh; and Wang, Peha, and Sirbu on integrated services.

[7] For a discussion about incomplete contract theory see Brynjolfsson, Erik, “An Incomplete Contracts Theory of Information, Technology and Organization,” Sloan School of Management, MIT, 1991; or Hart, Oliver, and John Moore, “Property Rights and the Nature of the Firm,” *Journal of Political Economy*, Vol. 98, No. 6, 1990, pp. 1119-1158. Both works build on transaction cost economics which is discussed in greater depth by Williamson, Oliver E., “Transaction-Cost Economics: The Governance of Contractual Relations,” *Journal of Law and Economics*, Vol. 22, October, 1979, pp. 233-261.

[8] Kavassalis, Lee, and Bailey, “Sustaining an Independent Bearer Service Market” (1997).

[9] A network externality is a benefit to incumbent users of a network as an additional customer joins the network. For example, a telephone system involving only one person has zero overall benefit since this one individual cannot call anyone else. If a second person joins the network, the first user will benefit since he or she now has someone to call. The network externality is positive when the benefit is positive, and negative as the additional user becomes a cost. For more information see Farrell and Saloner, “Competition, Compatibility and Standards: The Economics of Horses, Penguins and Lemmings,” in H.L. Gabel, ed., (1987); and Katz and Shapiro, “Network Externalities, Competition, and Compatibility,” (1985).

[10] The Internet “cloud” is the network of networks that comprises the Internet. It was drawn as a cloud as opposed to a specific diagram because typically one actually knew what was inside the cloud. Because the interconnection points and technology within the cloud are very dynamic and have distributed control, it is difficult to know what is inside. Traffic is injected at one point of the cloud and it exits from a different point of the cloud depending on the technology, interconnection points, and congestion at the time of transmission.

[11] There may be a benefit to having a smaller network of people to decrease the “noise” from the many users on a large network—or information overload, as it is sometimes called. However, we are focusing our discussion

---

on the Internet where users connect to every other user on the network, thus making them susceptible to this noise anyway.

[12] An example of this is the interconnection of two countries with the United States. Even though these countries may be adjacent and have similar technology, they both will benefit greatly by connecting to a network with greater technology (such as networks in the United States) than they do with each other. The effect this has is that countries that are very close route traffic through the United States and experience longer delay paths and cause more congestion in the U.S. portion of the Internet. For example, Australia and New Zealand had to traverse the U.S. backbone in the early 1990s to exchange traffic.

[13] Bohn, Braun, et al., "Mitigating the Coming Internet Crunch: Multiple Service Levels via Precedence" (1994); and MacKie-Mason and Varian, "Pricing the Internet" (1995).

[14] McKnight and Bailey, *Internet Economics* (1997).

[15] See Anania and Solomon, "Flat—The Minimalist Price," in McKnight and Bailey, ed., *Internet Economics* (1997) for an analysis of the benefits of flat-rate pricing for data networks such as the Internet.

[16] Edell, McKeown, et al., "Billing Users and Pricing for TCP" (1994).

[17] Brownlee, "New Zealand's Experiences with Network Traffic Charging," in McKnight and Bailey (1997).

[18] Kelly, Sharifi, and Petrazinni, *Challenges to the Network: Telecoms and the Internet*, (1997). Also partially from http://www.isi.edu/div7/ra/NAPs/naps_na.html and http://www.isi.edu/div7/naps/naps_eu.html.

[19] McKnight and Bailey, *Internet Economics* (1997).

[20] Evagora, "World Wide Weight" (1997); also at <http://www.teledotcom.com/0997/features/tdc0997globe.html>.

[21] "The European Internet: Enter the Telco," report by The Yankee Group, EuroScope Module, May 1997.

[22] Evagora, "World Wide Weight" (1997).

[23] Gillett and Kapor, "The Self-Governing Internet: Coordination by Design" (1997).

[24] Shenker, Service Models and Pricing Policies for Integrated Services Internet" (1993).

[25] Coase, "The Nature of the Firm" (1937); and Williamson, *Markets and Hierarchies: Analysis and Antitrust Implications* (1975).

[26] This analysis is discussed in detail in Bailey, "The Economics of Internet Interconnection Agreements" (1997). In summary, he finds that there is an incentive for customers to aggregate their traffic before reaching an interconnection point when a flat-rate pricing policy is used. Bailey also describes the ability of policing to eliminate aggregation of traffic or resale of the interconnection service but that, he argues, is very costly and reduces the trust between the interconnecting firms. Therefore, this article does not explore the ability to police for the enforcement of contracts. Furthermore, it should be added that the CIX tried to use a policing policy to prevent aggregation and resale with little success.

[27] The "tragedy of the commons" occurs when a public good, which can be shared among many people, is overused since each individual benefits more from more use of the public good. Eventually, the overuse of the good (or commons) leads to its eventual destruction. See Hardin, "The Tragedy of the Commons" (1968).

[28] MacKie-Mason and Varian, "Pricing the Internet" (1995).