Competitive Effects of Internet Peering Policies

by

Paul Milgrom, Bridger Mitchell and Padmanabhan Srinagesh

Reprinted from *The Internet Upheaval*, Ingo Vogelsang and Benjamin Compaine (eds), Cambridge: MIT Press (2000): 175-195.

ABSTRACT

This paper analyzes of two kinds of Internet interconnection arrangements: peering relationships between core Internet Service Providers (ISPs) and transit sales by core ISPs to other ISPs. Core backbone providers jointly produce an intermediate output -- full routing capability -- in an upstream market. All ISPs use this input to produce Internet-based services for end users in a downstream market. It is argued that a vertical market structure with relatively few core ISPs can be relatively efficient given the technological economies of scale and transaction costs arising from Internet addressing and routing. The analysis of costs identifies instances in which an incumbent core ISP's refusal to peer with a rival or potential rival might promote economic efficiency. A separate bargaining analysis of peering relationships identifies conditions under which a core ISP might be able to use its larger size and associated network effects to refuse to peer with a rival, thus raising its rival's costs and ultimately increasing prices to end users. An economic analysis of competitive harm arising from refusals to peer should consider cost-based, efficiency-enhancing justifications as well as attempts to raise rivals' costs.

Paul Milgrom
Department of Economics
Stanford University
Stanford, CA 94305-6072

Bridger Mitchell and Padmanabhan Srinagesh
Charles River Associates
285 Hamilton Avenue
Palo Alto, CA 94301

1 Introduction

This paper describes the technology and organization of Internet services markets and analyses how *peering arrangements* among *core* Internet Service Providers (ISPs) can affect efficiency and competition in these markets.

In the present market organization, a limited number of core ISPs exchange traffic with each other at private and public Internet exchanges. The arrangements among the core ISPs have two key features: each core ISP negotiates a separate interconnection arrangement with each other core ISP, and each accepts only traffic destined to one of its own customers. These peering arrangements result in the creation of *full routing tables*, which define the set of addresses that can be reached over the Internet. A much larger number of non-core ISPs enter into different arrangements for exchanging traffic, in which they purchase *transit* from core ISPs, who then accept traffic destined for any Internet address. Thus, the market for Internet services has a vertical structure in which core ISPs produce an intermediate output (full routing capability or core Internet service) that is used by them and by non-core ISPs to produce Internet services for end users.

A core ISP is one that maintains a full Internet routing table. To accomplish this, the core ISPs maintain peering relationships with all other core ISPs. When a core ISP refuses to renew an existing peering arrangement or establish a new one, anti-competitive behavior may be alleged by the refused ISP. This paper develops an analytical framework that can be used to evaluate such allegations.

In Section 2, we analyze the technology of packet routing for Internet services. Combining that analysis with a transaction costs analysis, we conclude in Section 3 that the cost-minimizing industry organization must consist essentially of a limited number of core ISPs who supply transit to a larger number of non-core ISPs. From this perspective, refusals to peer can sometimes be consistent with and even necessary for cost minimization in the provision of Internet services.

While Section 3 is devoted to an analysis of costs, Section 4 focuses on revenues to see whether direct bargaining among ISPs is likely to result in efficient peering relationships and competitive prices. This section assumes that all costs are zero. With zero costs, a *bill-and-keep* arrangement, in which neither core ISP pays the other for interconnection, is the cost-based, competitive benchmark. A simple bargaining model is used to determine sufficient conditions for bill-and-keep to be the equilibrium outcome when two core ISPs negotiate an interconnection arrangement. With realistic extensions of the simple model it is found that a large core ISP may have the incentive and ability to raise the price of peering to smaller core ISPs. Because costs are zero, refusals to peer on a bill-and-keep basis are anti-competitive, and, unlike the cost-based refusals of Section 2, do not increase economic efficiency.

In an antirust analysis of a refusal to peer, the central question is whether the refusal was a cost-based decision that may have served to maintain quality in the pool of core ISPs or an anti-competitive act that could raise prices paid by end users. A concluding section summarizes this paper's bearing on that question.

2 The Technology of Providing Internet Services

The Internet is an interconnected global network of computer networks based on the Internet Protocol (IP). The seamless interconnection of the Internet's constituent networks permits any subscriber to communicate with any other subscriber, regardless of the ISP from which the subscribers obtain their Internet connections. This seamless connectivity is the result of reciprocal and non-reciprocal interconnection arrangements negotiated by ISPs, and is strongly influenced by the capabilities and limitations of key Internet standards and technologies.[1]

To provide a basis for examining the structure of the interconnection agreements, we begin with a brief overview of Internet addresses and routing.[2] We then discuss factors that give rise to a hierarchy of core and non-core ISPs.

2.1 Internet Addressing

An IP address is a 32-bit string of zeros and ones. Traditionally, IP addresses belonged to one of three primary classes: A, B, and C. Class A addresses begin with a '0'; the first 8 bits (the net ID) identify the network, and the remaining 24 bits (the host ID)

---

[1] A concise summary of Internet history and technology can be found in Jeffrey K. Mackie-Mason and Hal R. Varian, "Economic FAQs About the Internet" in Lee W. McKnight and Joseph P. Bailey, *Internet Economics*, The MIT Press, Cambridge, 1997. A more detailed description can be found in Douglas E. Comer, *Internetworking with TCP/IP*, Volume 1, Prentice Hall, Upper Saddle River, 1995.

[2] Internet addressing and routing standards have been developed under the auspices of the Internet Engineering Task Force (IETF). These standards continue to evolve rapidly. Conclusions based on economic analyses of current Internet technologies may no longer be valid if the technologies change sufficiently.

identify the host. Class B addresses begin with ‘10’; they reserve 16 bits for the net ID, and 16 bits for the host ID. Class C addresses begin with a ‘110’, reserve 24 bits for the net ID, and 8 bits for the host ID. [3] Often, the net ID identifies the Local Area Network (LAN) to which the host computer is attached. Originally, the net ID was used to route packets between networks while the host id was used to route packets within a network.

The A-, B- and C-class addresses do not use the IP address space efficiently. Since administrators prefer to have a different net ID for each LAN they manage, and since many LANs need more than the 254 host addresses that are possible with a C-class address, higher-capacity addresses have been in great demand and the supply of Class B addresses has been rapidly depleted. To address this depletion, multiple class C addresses could have been assigned to administrators requesting a few hundred or a few thousand addresses. Since there are only 254 usable host addresses in a class C address, this approach would have resulted in a more efficient utilization of the address space than the assignment of class B (or A) addresses to relatively small networks. But this approach would have led to a larger number of net IDs, increasing the memory requirements of key Internet routers that maintain “full routes.” Continued proliferation of class C addresses (there are potentially more than 16 million) could have resulted in unwieldy routing tables that could not have been processed by available routers.

---

[3] Class D, beginning with 1110, and Class E beginning with 11110 were reserved for multicast and future use, respectively.

Classless InterDomain Routing (CIDR) was devised to use the address space more efficiently, while keeping routing tables manageable. Loosely speaking, a CIDR route is described by a 32 bit IP address and an associated 32 bit *mask* that consists of a sequence of '1's followed by a sequence of '0's. The '1's in the mask determines the network portion of the destination address. With CIDR, the original three address classes are expanded, since masks can vary in length from 8 bits to 24 bits, allowing twenty-four address "classes." Multiple class C addresses are not necessary for a network that may have a few thousand hosts: a single suitably-sized CIDR block (of addresses) will suffice. CIDR has been a useful compromise between using the address space efficiently and minimizing the size of key routing tables.[4]

2.2 Internet Routing

The Internet can be represented as a set of nodes interconnected by physical links such as private lines, Ethernet or FDDI buses, and Frame Relay or ATM Permanent Virtual Circuits (PVCs). Packet switches (called routers) at the Internet nodes switch packets in accordance with the Internet's routing protocols. When a router receives a packet, it reads the packet header to obtain the destination address, consults a routing table, and forwards the

[4] Early routing protocols (such as the Gateway to Gateway Protocol or GGP) that were used with *classful* addresses were designed without CIDR masks in mind, and could not accommodate CIDR. A new routing protocol, Border Gateway Protocol version 4 (BGP4), was required to implement CIDR in the Internet.

packet on the appropriate link to another router that is closer to the final destination.[5] With this "next hop" routing decision each packet is independently routed. Under normal circumstances, all packets in a message will follow the same path across the Internet. However, if a link or node failure occurs while a message is being transmitted, routers will update their tables to find alternate paths to the destination, and packets sent after the failure will take a different route from that of packets sent earlier. Computers at each end of the communication reconstitute the message from the packets received.

The acquisition of proper routes and the maintenance of accurate routing tables are critical functions of routing protocols and router management.[6] We focus on key differences between unsophisticated routers that use default routing and core routers capable of default-free or full routing.

2.2.1 Default Routing

Conceptually, a routing table consists of pairs of the form *(N,R)* where *N* is the IP address of the destination network (the net ID), and *R* is the IP address of the next router or next hop on the path to *N*. Routing tables range in size from a few entries to tens of thousands of entries. The size of a router's routing table is determined by the function served by that router. A router may have specific entries in its routing table only for hosts and

---

[5] Routers do more than forward packets: they compute checksums, fragment and re-assemble packets to conform with the requirements of the physical networks they ride on, enforce time-outs, and perform other functions as well. A more complete description of routing can be found in Comer, *op cit*, Chapters 8, 14-16.

[6] A comprehensive discussion of routing is well beyond the scope of this paper. For more detail, see Comer, *op cit*, Chapters 14-16.

routers to which it is directly linked. In this case, the router will forward a packet with a destination address that does not appear in the routing table to a *default* router. This simple routing scheme can be implemented with a relatively small, static routing table. The initial routing table can be entered manually during the installation of the router and modified manually when hosts and routers are added to, or dropped from, the physical networks to which the router is attached. The routing costs (including the cost of the physical routing device and the ongoing costs of maintaining routing tables) are relatively low.

If all routers were to use default routes, however, routing anomalies could arise. If two routers point their default routes at one another, a packet for a destination that does not appear in either router's routing table would loop back and forth until it was dropped, consuming router resources and resulting in greater congestion. There is also an increased likelihood that packets routed with partial information will follow longer paths than routers with complete information on all destinations.

2.2.2 Core Routers

ISPs have adopted a coherent and workable routing scheme by designating a set of *core* routers, none of which relies on default routes. Core routers typically process packets destined from one network to another, basing their routing decisions on the net ID of the destination. When a core router encounters an unknown destination address, it does not point to a default router, but drops the packet and returns an error message to the source.

The set of *full routes* contained in the core routers' tables defines the reach of the Internet. Each device associated with an IP address (or net ID) stored in the core routers' tables can communicate with all other devices associated with IP addresses in those tables. A device that is not associated with any IP address in the full routing table will be invisible to a large portion of the Internet. Given the rapid expansion of the Internet, maintenance of the core routing tables is a critical and demanding task for the core ISPs.

Border Gateway Protocol version 4 (BGP4) is used by core routers to develop their full routing tables. A greatly simplified description of this protocol is provided below. Core ISPs establish "peering sessions" between pairs of routers. When a peering session is initiated, each router *announces* (or sends to the other router) its initial routing table: the destinations or routes it can reach directly, the distance in hops to these routes (zero), and the router on the next hop (itself). When each router receives the other's routing table, it updates its own routing table to show the additional routes, noting that they are one hop away and that they can be reached through the other router. In case the two routers peer only with each other, a single update is sufficient to provide each router with a complete routing table. When a router peers with more than one other router, a more complex series of updates is necessary for the routers to converge to a consistent set of full routes. The routing tables are constantly updated as new customers with new addresses join the Internet, and as existing customers switch providers or drop off the Internet. The of BGP4 is its ability to automate the propagation of route changes throughout the Internet is attractive and dangerous.

2.2.3 Routing Among Core ISPs

Using BGP4, ISPs can engage in "policy routing": designating preferred routes, limiting the routes they announce to a peering router, and limiting route announcements that they will accept from a peering router. Policy routing provides some support for two interconnection arrangements prevalent in the U.S. Internet: transit arrangements and peering.

In *transit* arrangements, one ISP pays the other to provide connectivity to all addresses in the Internet. The ISP supplying transit accepts packets from the purchaser addressed to any Internet destination and delivers packets addressed to the purchaser from any Internet source. The purchasing ISP will typically configure its routers to point a default route at the transit supplier.

In *peering* relationships between two ISPs, each ISP accepts packets addressed to, and delivers packets from, its customers (i.e., the end users and the ISPs who purchase connectivity from it). Neither ISP points a default route at the other. ISPs in peering relationships typically "bill-and-keep" and do not make payments to one another.

In general, it is difficult and costly for a core ISP to ensure that its peers are adhering to the terms of the peering relationship and not surreptitiously pointing a default route at it. In practice, trust must often a substitute for verification. According to some ISPs, this trust is sometimes abused.[7]

---

[7] Recent discussion on an email list points out the need for, and availability of, prototype software that can be used by an ISP to detect whether another ISP has pointed a default route at it or used it for transit when it is not authorized to do so. See the thread: "Some abuse detection hacks …" at http://www.merit.edu/mail.archives/html/nanog/maillist. Also see *The Cook Report*, Gordon Cook, "Randy Bush on Technical Peering Issues", pages 9-13, available from http://www.cookreport.com.

2.2.4 Technical Efficiency of the Routing Hierarchy

The continued rapid growth of the Internet has generated an urgent need to adopt routing arrangements that economize on equipment investment, maintenance, and communications capacity and that are also flexible and responsive to changing circumstances. The system has evolved a relatively cost-efficient, hierarchical set of ISP relationships.

To appreciate the technical efficiency of the routing hierarchy, consider as a benchmark a fully meshed Internet in which every ISP is peered with every other ISP and no ISP purchases transit. Each ISP would need a physical link to every other ISP, and each ISP would need to manage at least one router with a full routing table. Most small ISPs do not possess the resources to order and manage such a large number of links, or to operate and maintain the complex routers required to obtain and update full routes. In addition, each ISP would need to establish and maintain technical and business relationships with every other ISP. With thousands of ISPs, the transaction costs of the fully meshed Internet would be prohibitive.

Instead of a fully meshed network, as described earlier a routing hierarchy has developed for the Internet in which a few core ISPs operate and manage default-free core routers while the remaining ISPs point default routes to one or more core ISPs. This form of hierarchical routing economizes on routing and transactions costs. The core ISPs typically peer with each other on a bill-and-keep basis; each core ISP bills its customers, keeps the revenues and exchanges packets without charge. Non-core ISPs typically purchase transit from a core ISP and point default routes at it.

3 Cost-Minimizing Organization

There are two approaches to the analysis of the way the organization of the Internet affects costs. The first is the "*technology-based*" approach. It focuses on technological issues, such as unnecessary duplication of equipment, hardware and software costs, the number and nature of interconnection points required at any time, and the costs of maintaining and operating networks. This approach takes as given existing relationships between technology and market organization it does not attempt to analyze and explain those relationships. In our technology analysis, for example, we take it as given that separate core ISPs own separate routers.[8]

The second is the "*transaction costs*" approach, which compares costs across various forms of organization, holding the technology constant. In its purest form, the transaction cost approach demands that claims about the effects of organization on technology be justified by comparing the transactional problems created by different forms of organization. For example, it may lead one to ask why there should be any difference at all between what

[8] This assumption was not satisfied by the first Internet interconnection arrangement linking commercial ISPs together. In that arrangement members of the Commercial Internet Exchange (CIX) exchanged packets through the CIX router, which was under the control of the CIX membership. The history of the CIX, which is not reviewed here, offers hints about the difficulties of shared ownership of routing equipment and the reasons why backbone providers now own and operate separate routers.

can be technically implemented by a single centralized system manager and what can be implemented by a set of independent core ISPs.

Our discussion in this section incorporates elements from both approaches. We consider economies of scale in coordinating and operating the system (technological considerations) as well as the need to minimize free-riding, overcome network externalities, and encourage coordination on standards (transaction cost considerations). We do not, however, attempt to explain the relationship between technology and market organization; rather, we treat the existing, consistent empirical pattern as an input to our analysis.

3.1 Economies of Scale

The cost of core routing, holding other factors constant, is likely to be sub-additive.[9] To see why, hold the number of end users and the volume of end user traffic constant. As the number of core ISPs increases, the number of core routers, each of which contains entries for *all* Internet routes, also increases. Moreover, more skilled personnel are necessary to maintain and manage core routers, as the maintenance of consistent routes is more difficult when there are more sources of route announcements, and more potential sources of error. Consequently, total industry-wide expenditure on core routing increases.

[9] Subadditive cost functions imply that one firm can produce any given level of industry output at lower cost than multiple firms can. See William W. Sharkey, *The Theory of Natural Monopoly*, Cambridge University Press, 1982, page 2. We assume in this analysis that interconnection between separate ISPs must respect existing negotiated standards and take place only at public interconnection points while interconnection among nodes of a single ISP can be engineered at the ISP's discretion.

The costs at public interconnection points also rise with the number of core ISPs. Each core ISP leases a link (such as a private line) from each public interconnection point to a nearby network node. A fully meshed network would require that regional ISPs establish nodes near public interexchanges or lease expensive long distance links from one of their network nodes to the public interconnection point. The multiplicity of long and relatively low bandwidth links would cost more than the relatively short, high-bandwidth links from a few core ISPs to a public interconnection point. In addition, since each core router must communicate its routes to every other core router, the number of routing messages exchanged increases with the number of core routers. The increased routing traffic occupies more capacity in the shared media over which the peering routers communicate and increases in usage-sensitive costs. Higher speed LANs may be needed at public interexchange points to accommodate a larger number of core ISPs.

If core ISPs seek to avoid the higher costs of shared interconnection media by using private interconnections, as they have done in the U.S., cost subadditivity emerges for a different reason. In this case, each core ISP must establish links between its core routers and core routers of other ISPs. The interconnection links are often private lines connecting the core routers. The larger the number of core ISPs, the larger the number of links and associated router interfaces required. The costs of these additional resources will raise unit costs, implying that the industry would have lower costs (other things equal) if it had fewer core ISPs. [10]

Limiting the number of core ISPs and core routers might enhance efficiency. Indeed, Comer concludes, “Core systems work best for internets that have a single, centrally managed backbone”.[11]

3.2 Coordination on Standards and Business Practices

As a network of networks, the Internet depends on coordination on for its success. The value of access to a given subscriber depends on the size of the whole network. Thus, the compatibility of, and interconnection among, networks increases the value of access to all subscribers.[12] However, compatibility and interconnection require coordination and cooperation particularly when technologies advance rapidly. Therefore, arguments based on network externalities might be used to justify coordination among ISPs. It is an accepted principle of transaction costs theory that coordination (and consensus) are easier to achieve

---

[10] Of course, a single company needs to connect its routers as well, but multiple core routers incur unnecessary duplication, for two reasons. First, unlike separate ISPs, a single ISP could consolidate its operations to avoid having two routers in locations in a single neighborhood. Second, even if two routers in one neighborhood were necessary, both need not be core routers. One could be a smaller router pointing a default route to the other (core) router. The savings in router management costs from having with a single router with an external BGP4 connection might be substantial.

[11] Comer, *op cit*, page 240.

[12] M.L. Katz and C. Shapiro, “Network Externalities, Competition, and Compatibility,” 75 *American Economic Review* 424 (1985).

when there are fewer parties involved.[13] In such circumstances, access restrictions might improve coordination on standards.

Addressing and/or routing standards are frequently updated to accommodate the rapid growth of the Internet. For example, as part of the transition from BGP3 to BGP4, UUNET (then Alternet), Ebone, ICM and Cisco established a virtual or "shadow" Internet for extensive experimentation. A crucial question was whether implementations of BGP4 could process classless (CIDR) routes efficiently.[14] When the protocol was found to work, it was implemented on other major backbones. It is doubtful that the rapid transition that was required could have been completed if the consensus of many providers had been required.[15]

IP addresses are currently defined by the IP version 4 (Ipv4) standard, which includes the original "classful" addresses and the extension to CIDR addresses. While CIDR has addressed some of the near-term problems of address exhaustion and routing table explosion, a newer standard, Ipv6, has been defined to accommodate the continued growth of the Internet. When Ipv4 addresses are eventually replaced by proposed Ipv6 addresses coordination among core ISPs can help minimize service disruption. Changes in other

[13] P.R. Milgrom and D.J. Roberts, "Bargaining Costs, Influence Costs and the Organization of Economic Activity," Perspectives on Positive Political Economy, edited by James E. Alt and Kenneth A. Shepsle, Cambridge: Cambridge University Press, 1990, 57-89. (Reprinted in Transaction Cost Economics, edited by Oliver Williamson and Scott Masten, London: Edward Elgar Publishing Co., 1994.)

[14] The CIDR standard had been designed to address the problems of address exhaustion and routing table exhaustion.

[15] "Experience with the BGP-4 protocol" by Paul Traina, RFC 1773, available at *http://www.internic.net*.

fundamental protocols will likely give rise to a need for further coordination, which may well be easier if there are fewer core ISPs involved. Ease of coordination might justify some refusals to peer.

3.3 Coordination on Router Management

Inefficient management of a core router by an ISP can impose significant external costs on other ISPs. Core routers exchange information with one another on how efficiently they can reach given destinations: in one hop, two hops, etc. Each core router uses the information received from other the core routers, combined with a measure of network distance, to build an efficient routing table that is consistent with the routing tables of other core routers.

When one core router sends incorrect information, *all* other core routers will build inaccurate routing tables, and *all* customers will experience degraded service (such as lost connectivity). Global problems can arise from local mistakes. *Route flapping* occurs when a router repeatedly announces and then withdraws a route. This initiates a series of upgrade messages which may cause routers to experience difficulty in converging on stable routing tables. The resources required to process these routing messages and compute routing tables can significantly limit the router’s ability to forward packets, degrading service to end users. A different global problem, *black-holing*, occurs when an ISP mistakenly announces a route that it cannot reach. Packets for the announced destination are sent to the ISP making the false announcement, but discarded because the ISP cannot reach the destination. Both these problems affect the services offered by the ISP and also the services offered by all other ISPs.

All ISPs have an incentive to correct these routing problems. ISPs with staff skilled in routing and router management are likely to solve a routing problem quickly. However, the solution will be of limited value unless *all* ISPs who exchange routing information implement it. Therefore, the solution must be shared with *all* core ISPs, including those with unskilled routing staff, or no routing staff whatsoever. However, this externality blunts the incentive to hire competent routing technicians, reducing the quality of service. The effects can be diminished by limiting membership in the core set to a limited number of ISPs with demonstrated routing expertise. Refusals by incumbents to peer with new entrants who lack the required routing skills can thus increase the overall efficiency of the Internet.

3.4 Free Riding on Backbones

The Internet routing architecture also provides incentives for ISPs to free-ride on their competitors' backbones. Consider a simple case in which two core ISPs interconnect on a bill and keep basis at two interconnection or interexchange points (IXs), one on the East Coast and one on the West Coast. Suppose that $host_1$, an East Coast customer of ISP_1, wishes to communicate with $host_2$, a West Coast customer of ISP_2. Rather than transporting packets from its customer across its backbone to the West Coast IX, ISP_1 would prefer to use the East Coast IX to exchange outgoing and incoming packets with ISP_2 so that transcontinental traffic would be transported on ISP_2's backbone. With bill-and-keep interconnection, ISP_1 has no incentive to transport the packets itself. ISP_2 has a similar incentive; it would prefer to use the West Coast IX to exchange incoming and outgoing packets for this customer. If one of the ISPs announced all of its routes at both IXs and the other announced only the local routes at each exchange, the latter ISP would free ride on the former ISP's backbone.

A core ISP might have to audit every other core ISP's route announcements to ensure that it was not being gamed in this fashion. However, the auditing, monitoring and enforcement costs arising from gaming may be quite high, and since monitoring tools are imperfect, some free-riding may occur even when monitoring tools are used.[16] The current practice of ISPs is asymmetric or "hot potato" routing, where each ISP delivers internetwork traffic to the other ISP at the IX nearest the source of the packet. With this compromise, each ISP uses its preferred interconnection point for traffic originated by its customers. When traffic flows are balanced, neither ISP takes a free ride on the other's backbone. By limiting interconnection to other ISPs with similar (uncongested) backbones or to ISPs that upgrade their backbones continually in response to increased traffic loads, an ISP can reduce the likelihood that its peers will seek to free ride on its backbone. Consequently, the ISP can economize on the costs of monitoring the interconnection agreements for compliance.

In sum, a hierarchical structure in which a few core ISPs peer with each other to produce full routing capability and supply transit services to a large number of other ISPs is likely to be more cost-efficient than a flatter structure in which a large number of ISPs peer

---

[16] "Randy Bush Discusses Technical & Economic Details of Peering & Routing Policy at Level of the Large Backbones", The COOK Report on Internet, November 1997 (Vol.6, No. 8), pages 9-13.

with one another to produce full routes. The socially optimal number of core ISPs is likely to be a relatively small subset of all ISPs, of which there are more than 5,000 in the U.S. Over time, the set of core ISPs may shrink, and the composition of the set may change as new entrants succeed in becoming core ISPs and others exit the market. In such a dynamic environment such as this, it is inevitable that some new ISPs may not be able to obtain peering arrangements with all incumbent core ISPs on satisfactory terms, and some core ISPs may not be able to renew all their peering arrangements. Some refusals to peer (with incumbents or with new entrants) might help maintain the economically efficient interconnection arrangements discussed in this section.

4 A Bargaining Approach to Peering Arrangements

Bargaining theory comes in two flavors: the older "cooperative bargaining theory" initiated by John Nash and the newer "non-cooperative bargaining theory" initiated by Jacob Stahl and rediscovered by Ariel Rubinstein. These two theories are closely connected: the Stahl-Rubinstein model duplicates the results of Nash bargaining theory under appropriate circumstances.

The older cooperative bargaining theory takes opportunity sets (that is, possible bargains) and threat points as its primitive elements. The threat points are conceptually problematic in the cooperative theory, for two reasons. First, the theory offers no way of assessing which threats are credible. Second, it it fails to distinguish between threat payoff that arises from taking outside opportunities to one from a temporary disagreement.

Non-cooperative bargaining theory is more specific, assigning different roles to outside options and costs incurred during disagreement.[17] It is also more flexible, offering a straightforward way to explore variations of the basic bargaining environment using the standard tools of non-cooperative game theory.

One of the greatest difficulties in applying bargaining theory comes from determining which threats are credible. Unfortunately, non-cooperative bargaining theory indicates that credibility is not merely a theoretical issue. For example, a labor union can decide whether its members will work during each period of a negotiation. Both sides incur large losses if the union strikes. By varying the parties' expectations about the conditions under which the union will strike or accept the firm's offer, Fernandez and Glaser show that there is a wide range of possible equilibrium outcomes.[18] Here, "equilibrium" means that (1) each party always acts optimally in its own interests, given its expectations and (2) its expectations are correct (in the sense that they correspond to the other party's planned behavior).

The existence of many equilibrium outcomes illustrates that bargaining outcomes are partly determined by history and culture, which affect expectations. To the extent that a core ISP with market power has exploited its position in the past by refusing to peer with a

[17] "The Nash Bargaining Solution in Economic Modelling" by Ken Binmore, Ariel Rubinstein, and Asher Wolinsky, Rand Journal of Economics, Volume 17, No. 2, Summer 1986, demonstrates the distinct roles and also the close connection between the new theory and the older Nash bargaining theory.

[18] Raquel Fernandez and Jacob Glazer, "Striking for a Bargain Between Two Completely Informed Agents," American Economic Review, 81, 240-252, 1991.

requesting ISP, such behavior should be expected in the future whenever it is consistent with that ISP's rational self-interest.

However, the threat of refusing to peer does not always empower a dominant network carrier to raise its interconnection price.[19] As the analysis below shows, other conditions are necessary for the exercise of market power even in a simple model.

4.1 A simple bargaining model

The following simple bargaining model can be used to analyze Internet peering arrangements. There are N homogeneous customers in the market, served by n core ISPs. Each customer obtains service from only one ISP and ISP_i serves a fraction α_i of the customers.[20] When ISP_i is not connected to any other ISP, its representative customer enjoys a benefit or utility of $u(\alpha_i, N)$ per period and is willing to pay a corresponding amount for that connectivity. Since we will be holding N fixed throughout this analysis, we use a less cumbersome notation by writing $f(\alpha)=u(\alpha,N)$ and conducting the analysis in terms of f.

Suppose that one core ISP serves a fraction α_1 of the customers and a second serves a fraction α_2, and that these proportions are independent of the interconnection arrangements

[19] An historical case in point is the arrangement reached in 19993 by ANS and the CIX. Although ANS, which operated the NSFNET, was the dominant carrier, it was forced to agree to the bill and keep arrangement proposed by the smaller CIX networks in order to provide the universal connectivity it had guaranteed to its customers.

[20] We assume that the customers of ISP_i include those who purchase service directly from ISP_i and also the customers of all non-core ISPs who purchase transit through ISP_i.

between the two ISPs. (This assumption implies the smaller ISP does not lose any customers to any other ISP when it loses connectivity to the larger ISP.) Suppose further that both ISPs have obtained peering arrangements with all the other ISPs. The revenues of ISP_1 would be $N\alpha_1 f(1-\alpha_2)$ if it did not obtain a peering arrangement with ISP_2, and $N\alpha_1 f(1)$ if it did obtain a peering arrangement. We assume for simplicity that there are no costs, so that revenues are equal to profits.[21]

Suppose the lack of interconnection is sustained only temporarily during bargaining, until the parties reach a peering agreement. The outcome of negotiations according to the non-cooperative theory (assuming identical discount rates for the two ISPs) is essentially the same as that of Nash bargaining theory with the no-interconnection payoffs as the threat point. The outcome is that the two parties divide the gains to cooperation equally, and the payoffs are:

$$\text{ISP}_1: \quad \pi_1 = \tfrac{1}{2}\{N\alpha_1[f(1-\alpha_2)+f(1)] + N\alpha_2[f(1)-f(1-\alpha_1)]\} \qquad (1)$$

$$\text{ISP}_2: \quad \pi_2 = \tfrac{1}{2}\{N\alpha_2[f(1-\alpha_1)+f(1)] + N\alpha_1[f(1)-f(1-\alpha_2)]\} \qquad (2)$$

With interconnection, ISP_1 will be able to charge each of its customers a subscription fee of $f(1)$, earning revenues (and profits) of $\pi_1 = N\alpha_1 f(1)$. In equilibrium, the payoff to ISP_1 is given by equation (1). The difference between the two payoffs is the negotiated net payment from ISP_1 to ISP_2. Since there are no costs in the formal model, such payments cannot be justified on the basis of costs imposed by one ISP on the other and thus a positive

---

[21] The implications of costs for peering arrangements were separately analyzed in Section 2.

net payment can be attributed to the exercise of market power. With some manipulation, the net payment can be shown to be:

$$N\alpha_1\{f(1) - f(1-\alpha_2)\} - N\alpha_2\{ f(1) - f(1-\alpha_1)\} \qquad (3)$$

The first term is the additional revenue that ISP_1 earns from its end users after it negotiates an interconnection arrangement with ISP_2. The second term is the corresponding expression for ISP_2. Thus, when both ISPs gain equally from interconnection, neither party pays the other, and a bill-and-keep arrangement is the equilibrium outcome of the bargaining process. When the parties do not gain equally from interconnection, the ISP that gains more will pay the other ISP for interconnection.

Sufficient conditions for bill-and-keep interconnection arrangements are easily obtained. If either (i) $\alpha_1=\alpha_2$ or (ii) f is linear $(f(\alpha)=a+b\alpha)$, then bill-and-keep is the outcome: $\pi_i = N\alpha_i f(1)$ for $i=1,2$ in equilibrium, and no net payments are made by either core ISP.

Indeed, the argument that with fixed numbers of customers the larger network has a general advantage in the bargaining depends on the shape of f. If ISP_1 is larger than ISP_2, then $\alpha_1 > \alpha_2$ and $f(1) - f(1-\alpha_2)$ is smaller than $f(1) - f(1-\alpha_1)$. The shape of f will determine whether ISP_1 pays or is paid for interconnection.

During early stages of market development when very few consumers have obtained Internet access, f may be almost linear so the simple model suggests that bill-and-keep arrangements should be relatively common. As market penetration increases, the value of connecting to additional subscribers may be subject to diminishing returns, and the relatively large core ISPs, whose own customers have a low marginal value of communicating with additional subscribers, may gain a bargaining advantage. This conclusion is consistent with the early history of Internet interconnection arrangements.[22]

4.2 Variations on the Simple Model

The preceding conclusion is derived jointly from the several assumptions of the model. Of critical importance is the assumption that customers are locked into a single network, i.e., that *switching costs* are prohibitive. Internet subscribers do face a range of switching costs when they shift from one provider to another. Large business customers are often required to relinquish their IP numbers and obtain new addresses from the range of CIDR blocks allocated to their new ISP. Renumbering can impose substantial costs on some business subscribers. Residential customers are often required to obtain new email addresses when they change ISPs, incurring expenses and inconvenience in the process. These costs of switching ISPs are similar to those incurred by telephone customers who change their telephone numbers when they switch Local Exchange Carriers (LECs). The switching costs for local telephony have been judged to be significant, and incumbent LECs in several countries are required by regulators to offer local number portability.

While Internet switching costs can be significant, the model's assumption that they are prohibitively high for all customers is extreme. This assumption is made operational in

---

[22] See "Internet cost structures and interconnection arrangements" by P. Srinagesh, in *Towards a Competitive Telecommunication Industry: Selected papers from the 1994 Telecommunications Policy Research Conference*, edited by Gerald W. Brock, Lawrence Erlbaum, Mahwah, New Jersey.

the simple model by assuming that α_i is independent of the number of subscribers that can be reached through ISP_i. If switching costs were low, the smaller network would lose at least some customers after being disconnected by the larger network and its profits at the threat point would be lower than they were in the simple model. The ability of subscribers to switch ISPs could weaken the bargaining position of the smaller network.

A second critical assumption is that there is only one source for the services provided by each ISP. A core ISP's ability to demand payment from another core ISP for connections to its customers depends on the absence of alternative routes to reach the same customers. Some downstream ISPs and large business customers purchase connectivity from two or more ISPs -- they *multihome*. Multihoming is technically complex and can be quite costly; only some customers are capable of taking advantage of the benefits it provides. However, the costs of multihoming are falling as new technologies such as Network Address Translation tools (NAT) are deployed.[23] Residential customers can similarly achieve a degree of independence by obtaining ISP-independent email accounts from providers such as Hotmail and Yahoo in addition to their ISP's email accounts. These forms of multihoming have an impact on the bargaining power of core ISPs. If all of the customers of one core ISP could be reached through other core ISPs, then that ISP's threat to withhold interconnection would not be credible.

---

[23] Praveen Akkiraju, Kevin Delgadillo and Yakov Rekhter, "Enabling Enterprise Multihoming with Cisco IOS Network Address Translation (NAT)", available at http://www.cisco.com/warp/public/cc/cisco/mkt/ios/nat/tech/emios_wp/htm.

A third assumption is that an agreement, once reached, is sustained indefinitely. The smaller core ISPs could have an incentive to merge to reduce their disadvantages in case their peering arrangements are threatened and so to increase their bargaining power. The Brokered Private Peering Group (BPPG) is one attempt at such a consolidation.[24] These firms may also have a dynamic incentive to expand to improve their bargaining position (though this must be balanced against the static incentive to shrink if there are increased variable connection costs). These growth incentives would inevitably cut into the larger core ISP's current profits. However, the large core ISP could not alter these incentives merely by forbearing from exercising its market power in the present, because a promise to continue forbearance is not credible. In this case, the desire to maintain a cooperative reputation is not likely to be an effective limit on the dominant ISP's behavior.

These considerations taken together suggest that, under some circumstances, large core ISPs may exercise market power in their negotiations with smaller ISPs by refusing to enter into, or to extend, peering arrangements with them. The emergence of a core ISP that is significantly larger than the others may harm competition. The larger ISP will have a bargaining advantage over its smaller rivals and can force them to pay interconnection fees that exceed the costs of interconnection. These fees may then have to be recovered by the smaller core ISPs through higher end user charges and by higher prices for transit charged to non-core ISPs.

---

[24] "Peering into the Future," by Randy Barrett, December 7, 1998. ISP Survival Guide, ZDNet.

One safeguard against this exercise of market power is vigorous competition among core ISPs. The simple bargaining model suggests that core ISPs of comparable size will enter into bill-and-keep arrangements, and no core ISP will then be able to raise its rivals' costs by raising the price of interconnection. At the same time, each core ISP will have an incentive to gain as many end users and non-core ISP customers as possible in order to maintain or improve its peering arrangements. Competition for these customers will tend to keep prices for transit and Internet service to end users low.

5 Conclusions

Our economic analysis of Internet interconnection concludes that routing costs are lower in a hierarchy in which a relatively small number of core ISPs interconnect with each other to provide full routing service to themselves and to non-core ISPs. Transaction costs analysis suggests that the market organization will mirror the routing hierarchy, as it does in current practice. In this hierarchy, refusals to enter into or renew peering arrangements can lead to lower routing costs and contribute to economic efficiency.

Routing costs, however, are not the whole story; account must also be taken of how peering decisions can affect the core ISPs' market power and consumer prices. A simple bargaining model of peering arrangements suggests that so long as there is a sufficient number of core ISPs of roughly comparable size that compete vigorously for market share in order to maintain their bill-and-keep interconnection arrangements, the prices of transit and Internet service to end users will be close to cost. If one core ISP can grow sufficiently larger than the others can and if customer switching during periods of disagreement is likely, then the bargaining model suggests that the largest core ISP can impose charges on other core ISPs. These charges will be in excess of the costs of interconnection (assumed to be zero in the formal model), and may over time strengthen the position of the dominant firm. The market may tip, and a single core ISP may dominate the upstream market for core connectivity.

An antitrust evaluation of a refusal to peer will, in general, need to consider both cost-based justifications of refusals to peer and allegations that a large ISP is exercising market power and harming competition. The cost-based justifications may be hard to quantify and the economic analysis of the competitiveness of peering arrangements is likely to be complex. This paper provides a framework for identifying, quantifying and integrating a range of factors that are important for such an antitrust analysis of peering.