THE CENTER FOR THE STUDY
OF INDUSTRIAL ORGANIZATION
AT NORTHWESTERN UNIVERSITY

Working Paper #0046

The Economic Geography of Internet Infrastructure in the United States

By

Shane M Greenstein*
Kellogg School of Management
Northwestern University

For the *Handbook of Telecommunications Economics, Volume II*
Edited by Martin Cave, Sumit Majumdar, Ingo Vogelsang

August 2004

---

*The Elinor and Wendell Hobbs Professor of Management and Strategy, Kellogg School of Management, Northwestern University. The author would like to thank Angelique Augereau, Howard Berkes, Oded Bizan, Pablo Boczkowski,Tim Bresnahan, Paul David, Barbara Dooley, Tom Downes, Chris Forman, Avi Goldfarb, Zvi Griliches, Brian Kahin, Mike Mazzeo, Sumit Majumdar, Eli Noam, Roger Noll, Robert Pindyck, Greg Rosston, Alicia Shems, Pablo Spiller, Dan Spulber, Scott Stern, Greg Stranger, Manuel Trajtenberg, Ingo Vogelsang, Joel Waldfogel, Glenn Woroch, and numerous seminar participants for many useful comments. Jeff Prince provided exceptional research assistance. All errors are the responsibility of the author.

Visit the CSIO website at: www.csio.econ.northwestern.edu.
E-mail us at: csio@northwestern.edu.

Abstract

In this review article I place industrial economics at the center of the explanation for the geographic properties of Internet infrastructure, applying this established mode of analysis to the unique cost structures and demand characteristics of these markets. I begin with an overview of the origins of the Internet and why the economic geography of Internet infrastructure was unclear prior to commercialization. The next sections analyze four overlapping markets of Internet infrastructure: dial-up service, backbone, broadband, and business provision of Internet infrastructure services. The concluding section summarizes the answers to several key questions motivating this broad research area.

Outline

1. Introduction

At some broad level, it is no great surprise that there is plenty of Internet infrastructure in Manhattan and not in the Mojave Desert. After all, the strongest factors that shape the location of demand for Internet services are the density of human settlement and the location of industry. Yet, that simple piece of economic reasoning only goes so far in addressing important questions about the location of Internet infrastructure.

In other areas of communications a combination of high sunk costs during installation and economies of density in operation make many areas very costly to serve. Remarkably, Internet infrastructure seems to defy that history. Less than a decade into commercialization, the supply of the most basic (i.e., dial-up) Internet infrastructure has become available nearly everywhere in the United States—except in the poorest rural locations. Was this achievement the intended or unintended consequence of government policy? Does that experience offer lessons about the strengths or weaknesses of relying on market forces to build out information infrastructure? These concerns are intertwined with a related event, the diffusion of the next generation of high-speed Internet infrastructure, generically called *broadband.* Unlike dial-up infrastructure, the earliest build-out of broadband has resulted in uneven geographic coverage that favors urban areas. Is this coverage an artifact of slow build-out or a permanent feature of relying on commercial markets for this service?

Addressing these questions presents quite a challenge. The Internet is unlike any communications network that came before it. It is a complex technology embedded in a multi-layered network, and many different participants operate its pieces. When the National Science Foundation (NSF) began to commercialize the Internet around 1992 there were no obvious precedents from which to forecast the Internet's geographic reach, nor were there simple lessons to borrow from other commercially supported communications networks.

In this review article I place industrial economics at the center of the explanation for the geographic properties of Internet infrastructure, applying this established mode of analysis to the unique cost structures and demand characteristics of these markets. It is challenging to do because Internet infrastructure involves an unprecedented mix of new economic actors and incumbent participants, something that one should expect of a new and large economic opportunity. However, unique and unprecedented market activity does not imply the presence of new economic precepts. Some basic (and often not mysterious) economic factors shaped the location of Internet Infrastructure.

This analysis makes reference to the same familiar economic concepts that have been seen with other communication networks. Especially central are economies of density and scale in operation, the economies of entry into services with high sunk costs, and the economics of retrofitting technical upgrades on existing infrastructure. The analysis also highlights the economics of competitive behavior for growing markets, behavior not unusual for many new technology markets, but somewhat novel for parts of communications. These themes are often neglected in common narratives about the Internet's geographic features.

Because the review also seeks to synthesize its themes with the already rich cross-disciplinary dialogue about the Internet's geography, the review presents its analysis at an intuitive level, eschewing formal economic theory, referring the reader to the writing of others when it is available.

I organize the narrative around six different topics. I begin with an overview of the origins of the Internet and why the economic geography of Internet infrastructure was unclear prior to commercialization. This first section highlights why the setting contributed to the Internet being both a surprise to most commercial participants, and, yet, also contributed to the Internet's fast diffusion. This analysis gives rise to themes that arise repeatedly later in the review, so it also serves as a foundation for analysis of the determinants of economic activity during the era of rapid geographic expansion.

The next sections analyze four overlapping markets of Internet infrastructure: dial-up service, backbone, broadband, and business provision of Internet infrastructure services. The section on dial-up focuses on why this mode of access became geographically ubiquitous so quickly. The discussion about broadband highlights why commercial activity in the backbone contributed to geographic ubiquity, and was not a hindrance to expansion, as some feared. The section on broadband highlights why broadband is comparatively more available in dense areas than areas of low density. The discussion about business infrastructure highlights progress and open questions in this comparatively newer area of research.

The concluding section summarizes the answers to several key questions motivating this broad research area. Why did near geographic ubiquity arise after commercialization? Why did market forces encourage extensive growth after commercialization? Has the Internet diffused disproportionately to urban areas? Is this Internet a substitute or a complement for urban agglomeration? Which policies mattered most and were these the intended or unintended consequences? What are the lessons for other countries?

2. Dispersion and Concentration of Internet Infrastructure

Before the Internet diffused there was no consensus about the shape it would take. To be sure, there was an active discussion among the telecommunications research and policy community about the economic geography of communications networks.[1] Even so, this only helped frame general issues, without providing sufficient precision to forecast specific outcomes. The presence of uncertainty is understandable, in retrospect. To oversimplify, new sets of economic actors developed new services, seeking demand for which there was only limited precedent. As with many new markets, participants had to experience a variety of activities to learn about key drivers of underlying costs, demand and operational efficiencies. It simply took a while to distinguish between long-term lessons and short-lived phenomena.

National and local policy during this era was also a mix of foresighted programs, which were then followed by reactive corrections to unanticipated issues. Some national and local policies anticipated some of the salient questions about the Internet's geography, because those questions arose as a by-product of concerns about regional comparative advantages, or about universal access to public communications networks. Even prior to the diffusion of the Internet, definitions for universal service had come under stress, as the diffusion of digital technologies and wireless communications highlighted regional disparities in access to frontier technology.[2] This experience helped to frame the first questions about the Internet. Only as the Internet grew much larger than anyone had forecast did it become apparent that policy makers needed to ask a new set of questions.

This topic provides useful background to the all the later sections, so it is natural to discuss first. The discussion focuses on three broad areas:

1. How could the Internet be a surprise and, yet, diffuse ubiquitously in only a few years after it became commercially available?
2. What were the origins of the commercial Internet and what aspects from this early era persisted into the commercial form? What changes caught most observers by surprise?
3. What pattern of geographic diffusion was first forecast and why? What issues did suppliers anticipate they would have to overcome to make the commercial Internet available outside a narrow set of big cities?

[1] For reviews, see, for example, Greenstein, Lizardo and Spiller (2003), Shampine (2001), Greenstein and Lizardo (1999), National Telecommunications and Information Administration (1995), and Kahin (1997).

[2] There are many places to enter the discussion about universal service issues in an era of technical change. See, for example, Crandall and Waverman (2000), Cherry, Hammond and Wildman (1999), Compaine (2001) and Noll and colleagues (2001), as well as many of the articles cited in this review. For new approaches more consistent with the concerns of urban geography, see Kotkin (2000), Castells (2002), Malecki (2002c), or Gorman and McIntee (forthcoming).

2.1 What is Internet Infrastructure?

In its most common usage, *Internet infrastructure* is synonymous with durable investments in software, communication and computing equipment, and related activities associated with operating information technology. This common and broad definition of Internet infrastructure encompasses quite a lot: capital equipment – such as mainframes, minicomputers, PCs (personal computers), LANs (local area networks), WANs (wide area networks), local and long-distance telephone equipment, private and quasi-public switching equipment, wireless networks for data transmission – and software – both packaged and customized. Notice that it also incorporates human capital, a key (and often local) input along any value chain for Internet services.

In addition, a proper economic accounting for the stock of information infrastructure encompasses the results of activities at user enterprises such as training, maintenance and operational experience. A proper economic accounting should also account for the stocks of knowledge and the flows of tacit and explicit knowledge about the technological frontier, which also often has local dimensions. [3]

During the early years of the commercialization of the Internet, the location of human capital was important. Universities had been the spawning ground for the Internet in two ways. First, the NSF subsidized the connection of many universities, which effectively trained many students in the basic operations. Second, these same subsidies encouraged many users to become familiar with the benefits of the technology's application – e.g., e-mail and file-transfers at one point followed later by the browser. This experience helped seed demand for commercial versions of the same service, particularly among recent university graduates.

It was often observed that the early commercial firms and users tended to be located near universities. Casually speaking, the high speed "pipe" was nearby, as were potential demanders and potential entrepreneurs to operate new businesses. As a result, there was an open research question at the outset of the Internet: As the commercial market for Internet services acquired its own market-oriented incentives, would this technology take the path of many new high-technology industries and concentrate within the location from which it grew? Or would the Internet divorce economic activity from the location of the major research universities? The answer turned out to be divorce with an exclamation point. The divorce occurred quickly.

2.2 The Origins of Internet Infrastructure

Nobody set out to design a commercial Internet with specific geographic features. These features arose endogenously, both borrowing from the geographic features of the network's origins and adopting new features to suit

[3] For a mildly different conceptual model and an attempt to trace the dollar flows of the internet value chain, see O'Donnell (2002). Another related model can be found in Gorman and Malecki (2002), who distinguish between the layers of the Internet affiliated with application and data transport, arguing that the geography of data transport affects the performance and value created in application.

market forces. The following overview of the early Internet helps explain why it was hard to forecast the geography of the commercial Internet from the features of the noncommercial Internet. Table 1 highlights some key events in the historical progression to the commercial Internet.

Even before its commercialization, Internet access was ultimately a local activity. In its dial-up form, the Internet employed a series of components, supplied by vendors and bought by users, retrofitted onto the U.S. voice network. That is, the Internet first grew on top of "plain old telephone service." Just as the user did not need the permission of the local telephone company to hook up a facsimile machine, the user of the Internet could hook his or her PC into the telephone system in order to download messages. This involved a local phone call to a local phone number, where a modem bank and server automatically handled the call.

Part of this was not an accident. Many advanced features of local telephony—such as clean lines free of static that made it feasible to send a digital signal—had been the focus of purposeful policies aimed at upgrading the U.S. telephone system. To be sure, many of these upgrades were done to enable services other than the Internet and the advance to digital technology in local telephone switches was not necessary for the diffusion of the first generation of the Internet, but their presence made later transitions to advanced technology almost incremental. Similarly, during the 1980s the U.S. telephone switch network made a technical leap to digital equipment from legacy analog equipment (Shampine, 2001). This leap spawned many new applications and services, everything from data-traffic to better faxes. Generally, there were many potential uses for digital switches, and the Internet was not the primary motvation at first.

Prior to 1992, the Internet consisted of the operations found at an academic modem pool or research center, with network interconnection handled by NSF-sponsored "regional" operations (Kahin, 1992). NSF's acceptable use policy prevented large scale commercial activity from occurring on the network. University centers were small-scale operations, typically serving no more than several hundred users, involving a mix of frontier and routine hardware and software. A small operation required: (1) a server to monitor traffic and act as a gate-keeper, (2) a router to direct traffic between the Internet and users at PCs within a LAN, and (3) an arrangement with a regional network, connecting the University to the Internet backbone or data exchange point. Revenues were not regularly collected in these arrangements and budgetary constraints were not representative of what might arise with commercial operations and competitive pressures.

The array of services matched the needs of academic or research computing, which had only a partial overlap with the needs of commercial users.[4]

[4] See Kahin (1992), Frazer (1995), Mowery and Simcoe (2002) or Kahn (1995) or Waldrop (2002) for a history of the development of the Internet under academic auspices and the policies supporting it, as well as how this translated into commercial development. Kende (2000) provides an excellent description of the commercial activities that aided the transition into the commercial era, such as the establishment of the Commercial Internet Exchange (CIX) and the privatization of NSF's Network Access Points (NAPS) and their eventual expansion. Note, too, university

Many small colleges had opened their Internet connections with NSF subsidies. A small staff of students or information technology professionals operated these centers. The organizational arrangement within research computing centers also was idiosyncratic, usually with only loose ties, if any, to the professionally run administrative computing centers of a university or research organization.

In all but the wealthiest universities, academic access centers operated in the cheapest way possible. They borrowed equipment from already existing facilities (i.e., PC and network servers) and charged little to students and staff. They borrowed practices common to bulletin board operations, and took advantage of local telephone pricing (i.e., by anticipating free local phone calls at no expense to users *and* by making no provision for long-distance calls or any other services). Points of presence (or POPs in industry parlance) began sprouting up in every university to support a geographically concentrated local user base. Although the primitive operations would eventually lend some aspects to the commercial operations, these university operations provided little basis for forecasting the form eventually taken by the commercial Internet.

2.3 Scale and Urban Location

In the beginning the Internet was a network for use by only researchers at laboratories, universities and military research centers. The Internet of 1992 was a world of insiders who treated it as a labor of love. The collection of insiders consisted of academic engineers who loved operating a network with frontier features, software gurus who loved experimenting across a wide array of multi-location applications, and policy experts who loved discussing policy issues for communications markets. Representative views from this group can be found in policy-oriented publications of the time (e.g., Kahin, 1992). These views display a mix of pride about accomplishments and competing visions for the future. One might also say that these insiders were focused on turning the Internet into a self-sustaining network (i.e., one that covered its own costs), but not necessarily a commercially successful network with appeal to the mass market.

Up until this point, the assets for operating the Internet were mostly in the public domain. The exception arose at private laboratories with government funding. In such a setting ownership of some Internet assets was private, but they operated in conjunction with the networks in the public sphere, such as the regional networks operated under the ultimate guidance of the NSF. After 1992, the Internet was headed towards an unknown future, where private firms could make all the investments in Internet infrastructure. The innocent past gave little hint about what would occur after the Internet commercialized. Nobody was openly predicting the type of wild growth that soon would take place.

Did anyone in 1992 forecast whether the Internet would agglomerate within major cities or not? The question is almost unfair in retrospect. As with many other government-subsidized technologies, it was unclear *whether* the Internet would grow at all outside of the quasi-public sphere that spawned it. The

---

access providers did not operate in a complete vacuum. The technology and operations in ISPs and bulletin boards had much in common. For this viewpoint, see editions of Boardwatch Magazine from the early 1990s, stored on www.boardwatch.com.

question of *where* it would grow was less paramount in 1992. To be sure, by 1992 investment decisions for the network backbone had been devolved to regional decentralized decision makers (Mandelbaum and Mandelbaum, 1992), but these arrangements came about as a result of NSF's guiding hands mixed with the cooperative nature of academic networking among insiders. A self-organized network operated by privately motivated participants was just starting at this point and – with NSF's blessing – could have gone in any number of directions.

Three considerations informed the discussion at the time about geographic diffusion. First, Internet infrastructure displayed economies of scale, but it was unclear how this would show up in a commercial setting. Second, the Internet was an advanced technology, and hard to support. Third, the policy environment at the time was in flux. This review addresses each of these in turn.

The technology of the Internet might have led to economies of scale, as found in the voice network. In voice networks there are economies of density in the provision of services linked to telephone switches and lines. This arises for two reasons. First, there are economies of scale in switching equipment. A large switch has a higher capacity per dollar than a smaller switch. Second, the lines for serving a denser area require less material and operational expense per customer.[5] It was possible that the building blocks of local Internet networks in the early 1990s – e.g., hubs and routers, Ethernet cards, and T-1 lines hooked into local area networks — might give rise to some form of density economies. Third, there were well-known forecasts that economies of scale in transmission technology would continue to increase, i.e., as fiber optic technology advanced and data backbone lines became capable of carrying more traffic.

Yet, in the early 90s it was not at all clear that these economies would continue to extend beyond the floor of an office or a large office building, as scale increased to city blocks, neighborhoods and regions. The NSFnet at the time was too special to give any hints about where or how it would deploy when the scale of use increased dramatically (See the discussion in Frazer, 1995 or, earlier, Kahin, 1992). If anything, it was mildly misleading to see it used as a network designed to facilitate sharing super-computing.

Moreover, Internet infrastructure initially appeared to be sensitive to a different geographic factor – the location of personnel to solve technically difficult problems. As with other advanced technologies, this one might favor urban areas where there are thicker labor markets for specialized engineering talent or supporting services. Similarly, close proximity to thick technical labor markets facilitates the development of complementary service markets for maintenance and engineering services.[6] Indeed, as commercialization got well underway, a pessimistic view emerged in the mid 1990s, one that became affiliated with the notion labeled "the digital divide." It forecast that the Internet was diffusing disproportionately to urban areas with their complementary technical and knowledge resources for supporting PCs and other computing, leaving rural

[5] See, for example, the discussion in Rosston and Wimmer (2001).
[6] These are closely related to the major reasons given for industrial agglomeration (Krugman 1991).

areas behind.[7]

There was only limited experience from which to forecast the importance of urban location. The NSF had deployed the Internet in a configuration that favored the needs of universities and researchers, with some emphasis on the needs of super-computer users in the late 1980s (Smarr and Catlett, 1992; Frazer, 1995; Kahn, 1995). This focus provided few general lessons. The self-contained operations found at university campuses or large laboratory settings provided no market model for the structure of profit-oriented operations, sales, and support to a large number of businesses and households. The commercial firms who first provided TCP/IP interconnection for a fee also faced large unknowns about the basic features of demand, such as its location, time-of-day properties, and so on.

Moreover, the NSF (following DARPA) had adopted an "end-to-end architecture," a feature that served academic needs by encouraging development of applications at the "ends of the system," where users could customize applications to their needs. That is, the applications ran in the PC or workstation client after data traveled across the Internet. Applications were disembodied from the specific location, the identity of the data carrier, or even the network configuration for moving data (Blumenthal and Clark, 2001). If location were not supposed to play a role in application development, how would a commercial deployment vary by location? It was not at all clear how a new set of customers would alter the deployment of Internet infrastructure across geographic space.

The commercial uncertainty was irreducible. It was unclear at the outset which of several potential maturation processes would occur after commercialization (e.g., for a standard discussion, see Rogers, 1995). If advancing Internet infrastructure stayed exotic and difficult to use, then its geographic distribution would depend on the location of the users most willing to pay for infrastructure. If advancing Internet infrastructure became less exotic to a greater number of users and vendors, then commercial maturation would produce geographic dispersion over time, away from the areas of early experimentation. Then, the diffusion would depend on the cost of building, supporting, and operating the network. Similarly, as advanced technology became more standardized, it would be also more easily serviced in outlying areas, again contributing to its geographic dispersion.

The other major source of uncertainty concerned the policy environment. It was not enough for the NSF to simply declare the Internet private, sit back and watch firms interconnect. It was well understood that several cooperation principles – such as voluntary interoperability among the components of infrastructure – kept the pre-commercial Internet operating. For example, all operators of components of the non-commercial infrastructure voluntarily interoperated with each other. What was to prevent private parties in the new era from choosing to not interconnect if it served a commercial purpose? Would the Internet balkanize from such action?

[7] Articulation of this view is particular prominent in the National Telecommunications and Information Administration's "Falling Through The Net" series (See National Telecommunications and Information Administration 1995, 1997).

There was simply no experience in the pre-commercialization period with uncoordinated commercial forces developing a communication network such as this. As a result, it was unclear what commercial governance structure was most appropriate for privately motivated investment between inter-connecting firms.[8] The Domain Name System also had to be transferred into the public domain, a process that turned out not to be smooth (Weinberg, 2002). The appropriate pricing structure for use of assets by private parties was also unknown.[9] There were open questions: Would the Internet require close regulation, as in the phone industry? The NSF put control over data-exchange points into regional hands, but would this cooperative structure be sustainable with infrequent guidance by government oversight?

In addition, there were national political pressures to reduce regional inequities in the deployment of the voice network, especially between urban and rural areas. It was not at all clear how, or whether, these policies applied to Internet infrastructure (For a full history, see Kahin, 1997). The uncertainty extended to the highest levels of the Federal government. Vice President Gore had supported federal subsidies for an "Information Superhighway", especially while it was primarily used for research. It was less obvious how this broad policy translated into specific laws after commercialization. For example, several federal agencies had historically pursued policies associated with mandates against regional inequality of access to advanced technology (e.g., they applied pressure for the adoption of digital switches in rural areas). It was unclear how to apply such principles to Internet access or broadband, if at all.

As it turned out, early into the commercial development of the Internet, the U.S. Congress passed the 1996 Telecom Act. This Act included provisions for the "E-rate program," which was aimed at alleviating geographic inequities in the provision of the Internet, among other goals.[10] It eventually raised over two billion dollars a year from long-distance telephone bills. This money was administered by the FCC, and primarily given to disadvantaged schools and libraries to develop Internet connections. In 1996, it appeared as if this might help those in isolated or rural locations. At its passage, however, it was not clear whether this government program would survive court challenges nor how it would shape the geographic reach of the Internet. This would take several years to figure out.

[8] Kende (1999) provides an excellent discussion about the emergence of several private institutions, such as private peering and privately operated public exchange points, that did not interfere with the operation of the commercial Internet.

[9] This situation gave rise to uncertainty in many aspects of the operations of the Internet. For a skeptical early discussion, see Mackie-Mason and Varian (1995). See Meeker and Dupuy (1996) for an early and comparatively sober assessment (that is, in comparison to later analyses) of the commercial prospects of many firms in the Internet infrastructure market. In some respects, the Internet's geographic reach seemed almost secondary to many contemporary policy observers, who were concerned about many seemingly more fundamental operational issues, such as covering costs and the scope of use. For further discussions of many of the policy and operational issues, see, e.g., Kahin (1992), Kahin (1997), Kahin and Nesson (1997), Kahin and Keller (1997), Kenney (2003).

[10] Due to its close identification with Vice President Gore, this provision of the Act was labeled the "Gore Tax" by opponents. To understand how this provision arose out of a longer history of policy initiatives about information technology in the US, see Kahin (1997).

Moreover, the 1996 Act altered the legal relationship between local telephone companies, who were usually local monopolies, and the competitive firms who interconnected with them. These provisions were quickly reviewed by the FCC and then challenged in court. These regulations had consequences for how broadband Internet reached businesses and homes, but any informed insider could easily understand that these regulations would not settle into a permanent set of rules for several years. It was another source of uncertainty.

2.4 Scale and Standardization

Internet technologies associated with textual information had incubated for twenty years prior to commercialization. Many of these building blocks of the mass market Internet were well past the degree of technical maturity necessary for mainstream use. Many were "standardized" by the early 1990s – in the sense of having refined interfaces that permitted interoperability with other pieces commonly found within homes and offices (see Kahin and Abbate, 1995). Several of these standards will be referred to throughout this narrative because they shaped the economic geography of the Internet – through bringing scale to new application development and scale to Internet users. That is, the technical problem faced by potential users of TCP/IP applications in different locations was *similar*.

Some of this was a propitious result of the fact that the most common computing configurations in the US were alike. Client server architecture in most business computing in the early 1990s involved a PC client, a network connection, and a server. Clients were largely IBM-compatible PCs with Intel-compatible chip sets. Apple and work station clients were in the minority. Network connections were Ethernet-compatible, with a few proprietary technologies – such as IBM's token ring – in minority use.[11] There was variety in servers. There were minicomputers, mainframes or workstations. The latter were mostly Unix boxes and most of these were compatible with the Internet and TCP/IP because such compatibility had been a procurement requirement for many years at DARPA, which bought many workstations throughout the 1980s. Finally, FCC mandates for physical connection resulted in similar plugs and sockets everywhere in the stock of communications customer premise equipment most commonly found at residences and businesses.[12]

Yet, many proprietary technologies seemingly stood in the way of the use of a network with non-proprietary standards, such as the Internet. Novell's Netware was widely used, but could be made compatible with TCP/IP messages only with great effort. Only later generations of Netware made the transition easier.[13] DOS users and generations of Windows prior to Windows 95 also

[11] The standardization on the descendents of the IBM PC by the early 1990s is widely documented. For an analysis, see, e.g., Bresnahan and Greenstein (1999). For an account of the standardization to Ethernet, see e.g., Von Burg (2001).

[12] See Kahin (1997), Oxman (1999), and Cannon (2001), for descriptions of the origins of these policies at the FCC and how these led to mandates over the equipment design of anything plugging into the US telephone system.

[13] For more on this, see, e.g., Forman (2002) for an account of business establishments' conversion to the Internet, or Kenney (2003) for an analysis of factors shaping the development

needed effort to download text messages (though, to be sure, any sophisticated user could do it). Once Windows 95 diffused, all TCP/IP programs became easier to use.[14]

Another key event was the end of IBM's opposition to converting its mainframes to server purposes using non-proprietary technologies in the mid 1990s.[15] With this decision one of the last major barriers to non-proprietary technology at large enterprises was gone. The remaining major barrier at large enterprises was Electronic Data Interchange (EDI) – an electronic standard for communicating transactions between firms. It had started to diffuse in the early 1990s and was famously complicated and difficult to use (and so, accordingly, only wealthy firms could afford to invest in it). Despite its drawbacks, it had high value to enterprises who had large numbers of regularized transactions, such as auto parts suppliers. The diffusion of EDI slowed considerably after the commercialization of the Internet. Yet, EDI applications were not immediately retired, because these investments were expensive, complex and essential to many on-going operations. More effort in the late 1990s went into making TCP/IP applications and existing EDI applications compatible.

Resistance to networking technology in the early 1990s was partly, but not primarily, associated with technical incompatibilities. Studies of adoption behavior (Bresnahan and Greenstein, 1997) associated resistance to networking with the costs of altering organizational functions and procedures. Organizational procedures only changed slowly. Any utopian forecast about getting a new network technology rapidly adopted at businesses would have had to face this rather sobering prospect. That said, there was no particular geographic pattern to this resistance.

As it turned out, by the mid 1990s, at many business establishments it was possible to use a non-proprietary standard, such as TCP/IP, as long as the ultimate purpose was simple inter-organizational communications. For simple applications, such as email or browsing, converting Netware was the largest barrier to overcome in the mid 1990s. Otherwise, most businesses could quite easily convert. For complex applications, however, many technical and organizational issues also mattered (Forman, 2002). At homes, the ownership of a PC and a modem also enabled conversion to use of the Internet for simple purposes.[16]

The similarities of the retrofitting problems across areas throughout the US supported economies of scale in the supply of standardized parts. Moving the

---

of Internet applications.

[14] Microsoft built TCP/IP compatibility into Windows in order to (1) foster diffusion of Windows NT as a server operating system, and (2) facilitate the use of TCP/IP programs on the client. This activity facilitated "plumbing", not applications. In this sense it helped the diffusion of the Internet after 1995. This should be distinguished from making a browser, which Microsoft did later. Netscape/Mosaic showed it could be very popular with users and profitable to sell. See, e.g., Cusumano and Yoffie (1998).

[15] This was a by-product of the strategy of IBM's new CEO, Lou Gerstner, to turn around the firm's performance by abandoning many proprietary legacy practices, instead focusing on developing integrated services for large scale enterprises, IBM's traditional customer base.

[16] For these points see National Telecommunications Information Administrations (1997), (1998) and especially (2000).

Internet into the mainstream commercial sector did not necessitate building a whole new Internet equipment industry, nor did it require the establishment of many local service, consulting, or maintenance firms. These were already there, supplying goods and services to the universities, business users, and home PC users. With a little modification these same firms could also supply Internet applications. Business users with investments in networking technology, such as LANs or simple client/server architectures, also could adopt basic features of the Internet with little further invention from their suppliers.

More to the point, when the Internet commercialized in the United States in 1992, there was no essential difference between the variety of computing found in New York City, Los Angeles, Omaha or anywhere else in the mainstream. The experience in different cities was similar enough that the availability of an upgrade to the Internet was available to everyone, albeit not in much the same way everywhere. The costs of the upgrade or retrofit differed between locations, but the solution to the engineering issue was sufficiently normalized and understood. In short, the Internet possessed the ability to generate scale economies in the sense that standardization of technical goods enabled “thicker markets” for engineering services to develop in many locations. The national IT consulting houses, such as Accenture or KPMG, also found it worthwhile to invest heavily in specialists with technical knowledge about how to apply TCP/IP across a variety of circumstances.[17]

The Internet achieved scale economies in another sense that had unforeseen dynamic consequences. The Internet allowed any user in any location within the US to communicate with any other who made the same upgrade to TCP/IP compatibility. This turned out to be important for building an installed base of applications and sharing them. It also helped get the Internet started in the first place. The commercial Internet did not start from scratch in 1992. Many applications had already started accumulating prior to commercialization. The Internet was attractive to use because it enabled communication with a large set of other computer users at other locations.[18]

That potential for a large scale computer-based communications network, in turn, looked inviting for vendors who would eventually start businesses affiliated with building Internet infrastructure. The unanswered question between 1992 and 95 was whether the benefits of scale economies would still motivate activity in many locations or just in a few homes and businesses of the *cognoscenti*. The earliest demonstration of the power of scale came in the early 1990s. When Tim Berners Lee developed html and the URL in a laboratory in Switzerland in 1989, it spread quickly throughout the computing world, the US included. When a team of University of Illinois undergraduates developed a browser called Mosaic in 1993, insiders thought it might be possible to diffuse it to non-technical users. Mosaic set records for how fast it spread among students

[17] The economics of standards has long emphasized how unified technologies can support large market activities. See David and Greenstein (1990) for a review of this literature. See Kahin and Abbate (1995) for the state of discussion at the birth of the commercial Internet. See Rohlfs (2001) for a general treatment of the issues and a particularly cogent application to the development of the Internet (Ch 13).

[18] By 1992, one millions of hosts existed.

and non-students alike.

Demand-side scale economies arose because the Internet appealed to more than technical users. It also appealed to a typical non-technical user. The basis for standardized mass market applications was email, instant message, and eventually, browsing. Market demand for simple uses arose almost everywhere and induced entry from new entrepreneurs taking advantage of the growing network (See Kenney (2003) for more analysis of these events). This demand was so large that it overcame the forces that might have otherwise confined the Internet to a few locations.

2.5 Summary

At a general level, the location of commercial Internet infrastructure had straightforward determinants. Like all information infrastructure, it was devoted to providing services to users in particular places. It was a local economic good because the service could only be delivered if it were supported by appropriate physical and human capital. These inputs were in a particular place, which shaped the costs and quality of the services received.

Yet, such a general description only goes so far. It overlooks the specific forces that shaped the diffusion of Internet infrastructure in the United States after commercialization. For example, what precise role would economies of scale or economies of density play in the availability of the Internet to urban and rural customers? The general description also does not resolve the uncertainty about competing economic factors. What organizational form would commercialize this technology and, once it became deployed on a national scale, would it resemble the organizations found inside the research universities? These were among the many open questions at the time of commercialization.

3. The Spread of Commercial Internet Access.[19]

I begin by discussing the spread of commercial Internet access providers, which subsequently became known as ISPs. I will primarily focus on dial-up providers, deferring discussion about backbone and broadband providers to a later chapter.

The deployment of Internet as a mass-market service depended on the commercial prospects of dial-up firms. Commercial vendors of Internet access first began sprouting in the early 1990s. Their growth accelerated in 1995 and exploded thereafter. In the latter half of the 1990s had three remarkable features:

1. They grew rapidly, inducing tens of thousands of new businesses to form around Internet technology. This infrastructure, in turn, attracted the creation of a whole value chain of support activities—such as hosting services, millions of web pages and business applications—thereby achieving the trappings of mass-market status.
2. The Internet supplier market spread rapidly: The Internet quickly became nearly pervasive across locations, which attracted considerable attention because it was a unique historical achievement. Rarely do new commercial technologies spread this rapidly.
3. ISPs offered a wide variety of business services. No consensus arose about the optimal combination of service lines to carry, which is indicative of uncertainty about the appropriate business model for commercializing the service.

3.1. The Activity of Commercial Suppliers

Internet Service Providers requires a physical connection because the architecture of the Internet necessitates this physical connection. Both under the academic and commercial network, as shown in Figure 1, the structure of the Internet is organized as a hierarchical tree. Each layer of connectivity is dependent on a layer one level above it. The connection from a computer to the Internet reaches back through the ISP to the major backbone providers. The lowest level of the Internet is the customer's computer or network.

These are connected to the Internet through an ISP. An ISP will maintain their own sub-network, connecting their POP's and servers with IP networks. These local access providers get their connectivity to the wider Internet from other providers further upstream, either regional or national ISP's. Regional networks connect directly to the national backbone providers. Private backbone providers connect to public (government) backbones at network access points.

The basic commercial transaction for dial-up Internet access and data transport did not raise prohibitive technical issues. Most often it involved repetitious and on-going transactions between vendor and user. A singular transaction arose when the vendor performed one activity, setting up Internet

[19] This section substantially borrows from perspectives in Greenstein (2001). See, also, Coffman and Odlyzko (2001).

access or attaching Internet access to an existing computing network.

If the ISP also operated the access for the user, then this on-going operation provided frequent contact between the user and vendor and frequent opportunity for the vendor to change the delivery of services in response to changes in technology and user needs. This worked well for users because in many cases an ISP was better educated about the technological capabilities than the user. In effect, the ISP sold that general knowledge to the user in some form that customized it to the particular needs and requirements of the user. At its simplest level, this provided users with their first exposure to a new technological possibility while educating them about its potential. Within a few years, little or no customization was required for most households. [20]

For example, these processes could be seen in the simplest details an ISP would set up for a new user. The most predominant means of communications is email. The email equivalent of bulk mail is called a listserv, where messages are distributed to a wide audience of subscribers. These listservs are a form of conferencing that is based around a topic or theme. Usenet or newsgroups are the Internet equivalent of bulletin board discussion groups or forums. Messages are posted for all to see and readers can respond or continue the conversation with additional postings. Chat rooms and IRC serve as a forum for real-time chat. The ISP could help the user learn to use such facets of the Internet.

Such help became a facet of the Internet access business. For example, building on its previous dial-up business, America On-line developed comparatively simple instructions for its users. It also developed on-line help that new users valued. Famously, it helped 'Instant-messaging' gain popularity among mass market users – even though the basic idea is quite old in computing science and equivalent functions even existed on the Internet in other forms.

ISPs also helped users gain familiarity with other basic tools of the Internet. Some tools have been supplanted, but the most common are WWW browsers, gopher, telnet, ftp, archie, and wais. Browsers and content have grown in sophistication from the interface developed by Tim Berners Lee. Similar remarks could be made as new functions came on-line, such as news and entertainment, ecommerce, application hosting, videoconferencing, and so on.

Especially with a business user, access often included more than exposure to the Internet, by providing the installation, maintenance and training, as well as application development. These types of transfers of knowledge and extensions of service typically involved a great deal of nuance, often escaped attention, and yet, were essential to developing infrastructure markets as an ongoing and valuable economic activity. The technology lent itself to this small-scale customization because it was easy to adapt to PC use or available LAN technology. The basic technical know-how did not differ greatly from routine knowledge found in the computing services sector prior to commercialization.

Because these general technical skills were widely available as parts of

[20] Technology transfers varied from the complex to the simple. The complex could involve issues, such as whether the user wanted to access company servers remotely. The simple could involve tailoring access to the generation of the PC operating system owned by the user.

other existing computer and communications technologies, the knowledge necessary for applying it to the Internet quickly became widely available in all major urban areas (Greenstein 2001, Mowery and Simcoe 2002)). The only locations lacking in supply were those with quite deficient support structures for a local PC service market. Only the poorest urban areas or smallest and most remote cities were deficient in this way. And even some of them – though not all – did develop adequate commercial support for such services within a few years.

In other words, at the outset of commercialization there was no doubt that a sophisticated and wealthy user could get Internet access in the Washington DC area or the San Francisco area or, for that matter, in any major city where an early leader in providing service, IBM GlobalNet, choose to locate a point of presence. The open question concerned much smaller cities who are comparatively isolated, cities under a half a million population, but as big as places such as Peoria, Illinois or Reno, Nevada.

As will turn out, the maturity and standardization of the technology largely did not become geographically concentrated in a few places. For all intent and purposes, technological maturity did not become a barrier to the spread of basic access after 1995. As I will discuss below, these themes manifest in myriad ways in the dial-up market in the mid to late 1990s.

3.2 Government Policy Encouraged a Diverse Geographic Supply

Some NSF decisions and FCC regulatory decisions aided the geographic dispersion of the commercial industry. When the NSF took over stewardship of the Internet backbone, it invested in developing a system of address tables and address systems (for routing TCP/IP based messages) that could be scaled. Data exchange points remained organized around the cooperative engineering principles used within the NSF days, favoring no particular location, nor any particular participant in the operating network. As a technical matter, interconnection with the public switch network did not pose any significant engineering challenges for dial-up firms.[21]

The NSF had encouraged uncoordinated commercial development through the absence of any directional edicts. Accordingly, Internet access evolved in an extremely decentralized market environment. Aside from the loosely coordinated use of a few de facto standards (such as the World Wide Web consortium, see e.g, Kahn, 1995, or Kahin and Abbate, 1995), government mandates after commercialization were fairly minimal. The ISPs had little guidance or restrictions and were therefore able to both tailor their offerings to local market conditions and follow entrepreneurial hunches about growing demand.

Furthermore, decades of debate in telephony had already clarified many regulatory rules for interconnection with the public switch network, thereby eliminating some potential local delays in implementing this technology on a small scale. By treating ISPs as an enhanced service and not as competitive

[21] This is a mild simplification. Vendors did have to learn the boundaries of what was possible, but the engineering challenges turned out to be comparatively small. See Werbach (1997) or Oxman (1999).

telephone companies, the FCC did not pass on access charges to them, which effectively made it cheaper and administratively easier to be an ISP (Oxman, 1999, Canon 2001). The FCC's decision was made many years earlier for many reasons and extended to ISPs in the early 1990s with little notice at first, since most insiders did not anticipate the extent of the growth that would arise by the the end of the 1990s.[22] As ISPs grew in geographic coverage and revenues and threatened to become competitive voice carriers, these interconnection regulations came under more scrutiny (Sidak and Spulber 1998, Weinberg 1999). I say more about this below.

In summary, a lesson for all insiders became widely understood by 1996-97: technical issues associated with building and operating ISPs were not difficult to solve. As a result, commercial factors, not the distribution of technical knowledge among providers, largely determined the variance of development of the basic dial-up access market within a few years after commercialization.

3.3 The Founding of Commercial Organizations

Shortly after the Internet's commercialization, only a few commercial enterprises offered national dial-up networks with Internet access, and they mostly targeted the major urban areas. Pricing varied widely (*Boardwatch Magazine*[23]). Most of these ISPs were devoted to recreating the type of network found in academic settings or modifying a commercial bulletin board with the addition of backbone connections, so interconnection among these firms did not raise insoluble contracting or governance problems. These ISPs also were devoted primarily to dial-up; some ISPs attempted sophisticated data transport over higher-speed lines, where the regulatory issues could be more complex and where competitive local exchange competitors were developing the nascent market.

Very quickly ISPs learned that low-cost delivery required locating access facilities close to customers because of telephony pricing policies across the United States. The US telephone system has one pervasive feature; distance-sensitive pricing at the local level, but no charge for extremely short distances. In virtually every state of the country, telephone calls over significant

[22] The Federal Communication Commission (FCC) first classified ISPs as enhanced service, similar to other on line services, such as bulletin boards. This decision built on top of many others, developed over several decades, designed to allow interconnection to the local telephone network by providers of customer premise equipment and enhanced service providers. This is a long story, often told triumphantly, sometimes as if the effects on the Internet were the intended consequence. See, e.g., Kahin (1997), Werbach (1997), Oxman (1999), and Cannon (2001). It is fair to say that when these rules were first developed, nobody was doing it in order to encourage the Internet. Their extension to the new commercial firms called ISPs was a natural extension of the treatment of bulletin board providers, who were a considerably smaller economic force in the 1980s than ISPs became in the 1990s. Other rules could have discouraged ISP entry (e.g., by requiring different ISPs to pay to universal service funds on top of subscriber line or business line charges) or imposed a different cost structure (e.g., per-minute charges).

[23] The earliest advertisements for ISPs in *Boardwatch Magazine* appear in late 1993, growing slowly until mid-1995, at which point *Boardwatch* began to organize their presentation of pricing and basic offerings. There was an explosion of entry in 1995, with thousands being present for the next few years.

distances (i.e., more than thirty miles) engender per-minute expenses, whereas local calls for very short distances are usually free. Hence, Internet access providers had a strong interest in reducing expenses to users by providing local coverage. Indeed, it is probably more accurate to say that the notion of locating in such a way to take advantage of free pricing had already been understood in the bulletin board industry, so it was not hard to understand or reapply to ISPs.

Access over dial-up lent itself to small-scale commercial implementations. Several hundred customers could generate enough revenue (at the basic unlimited access price of twenty dollars per month) to support physical facilities and a high-speed backbone connection in one location, so scale economies were not very binding. The marginal costs of providing dial-up services were low and the marginal costs of expansion also fell quickly, as remote monitoring technology made it inexpensive to open remote facilities. The marginal costs to users of dial-up service were also low in response, involving only incremental changes for organizations that had experience with PC use or LAN technology. It was easy to generate revenue in subscription models, where a commercial firm withheld availability of access unless payment was made. Hence, the economic thresholds for commercial dial-up service turned out to be feasible on a very small scale, which was encouraging to small firms and independent ISPs.

Many firms, such as IBM, AT&T and AOL, wrote up and implemented plans to access businesses and home users on a national scale, but the economic advantage of large branded identities did not preclude the entry of small-scale firms, at least not at first. Contemporaries noted that the largest firms played an important role as “certifiers” during the early years. This observation was frequently made about IBM’s provision to business and AT&T’s entry, which came sooner than any other large established firms’ entry into provision for homes. These actions helped to ratify and legitimize the market for business and home Internet access (Maloff, 1997). The on-line service providers – Prodigy, Genie, CompuServe, MSN, and AOL – also began converting to Internet service around 1995 too, with some providing service earlier than others.[24]

The failure of known firms to dominate that market (or induce a quick shakeout) was interpreted by many observers as a sign that small-scale firms had a viable opportunity to provide this new service as long as demand continued to grow (Greenstein 2001). Indeed, thousands of small entrepreneurial ventures also grew throughout the country and gained enough market share to sustain themselves. New entrants, such as Erols, Earthlink, Mindspring, Main One, Verio, and many others, gained large market positions. Some of these positions were sustained and others were not. Private label ISPs also emerged when associations and affiliation groups offered re-branded internet access to their members. These groups did not own or operate an ISP, instead their

[24] These conversions received considerable publicity at the time. For example, AOL’s struggles with the competition from the Internet received the much media attention. Its full conversion to the Internet came comparatively later than others. It finally adopted unlimited access to AOL content and Internet content in November 1996 – i.e., converting from “pricing per service” and “pricing per minute” to so called “all you can eat pricing.” The large user response overwhelmed AOL’s facilities. It took them several months to increase their modem/server capacity to handle the large volumes of calls and longer session lengths.

access was being repackaged from the original ISP and re-branded by the association. These firms could survive on relatively low market shares, though, to be sure, they were not very profitable either. By the end of the 1990s AOL's standardized service came to dominate market share, but that still left tens of millions of potential users for others.

The last successful mass market entrant were the "free" ISPs. The free-ISP model emerged in late 1998 and grew rapidly in 1999, offering free Internet access in exchange for advertisements placed on the users' screen. These firms eventually signed up several million households. Netzero became the largest of these. It converted to a hybrid free/charge model after its merger with Juno Online in 2001.

A few statistics will illustrate the trends. In one of the earliest Internet "handbooks," Krol (1992) lists 45 North American providers (8 have national presence). In the second edition of the same book, Krol (1994) lists 86 North American providers (10 have national presence). Marine, Neou and Ward (1993) lists 28 North American ISP's and 6 foreign ISP's, which is also about the first time *Boardwatch Magazine* begins to carry priced advertisements for ISPs. Meeker and Dupuy (1996) reports that there are over 3000 ISP's, and the Fall 1996 *Boardwatch Magazine's Directory of Internet Service Providers* lists 2934 firms in North America. By January 1998 the same magazine lists 4167 ISPs. Of course, the growth occurred due to the entry of small ISPs, each of whom had a tiny fraction of the market.

Another statistic also represents this growth. When Downes and Greenstein (2002) did their first survey of dial-up Internet access, they compiled a list of 12,000 local phone numbers for calling the Internet in the fall of 1996. By the Fall of 1998, when they did their last compilation, the whole list had swollen to over 65,000 local phone numbers. By that latter date every metropolitan area in the US had competitive coverage by then, as did many small towns.

3.4 Coverage By Dial-Up Providers

Growth and entry brought about extraordinary results. Downes and Greenstein (1998, 2002) have constructed statistics about the density of location of ISPs at the county level for the fall of 1996 and 1998.[25] See Figures 2 and 3 for illustration of ISP entry in 1996 and 1998.

ISPs tend to locate in all the major population centers, but there are also some providers in rural areas.[26] Their findings also illustrate the importance of changes over time. Many of the areas that had no coverage in the fall of 1996 were covered by the fall of 1998. Many of the areas that had competitive access markets in the early period were extraordinarily competitive in the latter period. This pattern seems to follow a pattern found in other new technologies that increasingly standardize over time, from urban areas into rural ones (Rogers

[25] For further documentation of these methods, see Downes and Greenstein (2002) or Greenstein (2001).

[26] In a light moment this pattern motivated (the late) Zvi Griliches – known for his research on the diffusion of new strains of corn – to observe that the map from 1996 showed that a good predictor for the absence of Internet Service Providers was the presence of hybrid corn seed.

1995).

Downes and Greenstein (2002) show that more than ninety-two percent of the U.S. population had access by a short local phone call to seven or more ISPs by 1998. Less than five percent did not have any access. Almost certainly the true percentage of the population without access to a competitive dial-up market is much lower than five percent. In other words, with the notable exception of some low-density areas, ISP service was quickly available almost everywhere.

The lowest density and poorest areas of the United States did lag, however, bearing signs of permanent retardation of development. In a study of the Appalachians and some areas with histories of poor communications service, Strover, Oden and Inagaki (2002) examine ISP presence in the states of Iowa, Texas, Louisiana, and West Virginia and determine the availability and nature of Internet services from ISPs for each county. They find that the rural areas suffer significant disadvantages for Internet service. Such counties often lack competitive supply of providers. Even when suppliers exist, they sometimes provide limited services or focus on specific segments, such as business users, not households. It is difficult for any policy – other than subsidy – to overcome these basic economic constraints.[27]

Strover (2001) arrives at a comparatively pessimistic assessment, one shared by many observers.[28] She points out that the cost structure for ISPs is unfavorable because of their dependence on commercial telecommunications infrastructure providers, which are reluctant to invest in rural areas due to the high costs necessary to reach what often are relatively few customers. Furthermore, a lack of competition in rural areas among telephone service providers serves to exacerbate the low incentives.

Conversely, there is an optimistic element to ISP rural service, affiliated with the dispersal of E-rate funds. As was previously noted, E-rate funds were dispersed in order to alleviate geographic inequalities in the dispersion of the Internet. Hence some researchers examined the provision of the Internet, not for the home or office, but for libraries and schools. For example, Bertot and McClure (2000) show that library use of the Internet is nearly ubiquitous, irrespective of location.[29] In their survey, the key issues for disadvantageously located libraries revolved around the quality of Internet activity and supporting services. Access was taken for granted nearly everywhere by the late 1990s.

Goolsbee and Guryan (2002) reach a similarly optimistic assessment based on their study of the California schools' access to the Internet. They too show that the discussion had moved to focusing on issues well beyond access. Again, they attribute this to the dispersion of E-rate funds in the late 1990s.

---

[27] For example, see Nicholas's (2000) study of the multiple attempts to provide access to rural Texas communities. He shows how the construction of calling area geographic boundaries shapes the entry patterns of ISPs. His close study of regulatory policy in rural Texas shows both the strengths and pitfalls of this policy approach.

[28] See also Garcia (1996), Gillett (2000), Hindman (2000), Parker (2000) or Malecki (2003).

[29] Indeed, earlier reports show that this had been true since the mid 1990s for a high fraction of libraries. See Bertot and McClure (2000) and further studies at http://www.nclis.gov/statsurv/statsurv.cfm for information on the near ubiquity of Internet access in US libraries by 2000.

3.5 Organizational Strategy and Coverage

An unexpected pattern accompanied this rapid growth in geographic coverage. First, the number of firms maintaining national and regional networks increased between 1996 and 1998. In 1996, most of the national firms were recognizable—they were such firms as IBM, AT&T, and other established firms that entered the ISP business to provide a complementary part to their existing services (e.g., providing data services to large corporate clients). In 1996 the other recognizable firms were those who had supported national on-line consumer services, such as CompuServe, AOL and Prodigy. All were in the process of converting their on-line service, previously run more like bulletin boards than ISPs, into Internet providers. By 1998, thousands of entrepreneurial firms maintained national networks and few of these new firms were recognizable to anyone other than an industry expert.

There was also a clear dichotomy between the growth patterns of entrepreneurial firms that became national and regional firms. National firms grew geographically by starting with major cities across the country and then progressively moving to cities of smaller populations. Firms with a regional focus grew into geographically contiguous areas, seemingly irrespective of urban or rural features. The growth of standardized contracting practices for renting facilities from other firms in other locations further added to the geographic reach of individual ISPs, even in areas where they did not operate or build facilities.

Most of the coverage in rural areas comes from local firms. In 1996, the providers in rural counties with under 50,000 population were overwhelmingly local or regional. Only in populations of 50,000 or above, do national firms begin to appear. In fall of 1998, the equivalent figures were 30,000 or lower, which indicates that some national firms had moved into slightly smaller areas and less dense geographic locations (Downes and Greenstein, 2002). Figures 2 and 3, taken from Downes and Greenstein provide a visual sense of those patterns in 1996 and 1998.

Local and regional ISPs provided Internet access everywhere, including small rural towns. National firms started with the big cities and slowly expanding into less dense territory. It appears as if it did not pay for many large national providers to provide dial-up service for the rural areas, whereas many small local firms in other lines of business (e.g., local PC retailing) could afford to add Internet access to their existing business. It may also be the case that the local firm may have had an easier time customizing the Internet access business to the unique needs of a set of users in a rural setting.

By 2000, many of these trends came to a halt. Growth in demand slowed, as Internet adoption at households surpassed 50% and was approaching 60% (National Telecommunications and Information Association, 2002).[30] Business adoption of basic Internet access for purposes of participating in email and browsing had also reached near saturation (Forman, Goldfarb, and Greenstein

[30] Over 65% of the US population made use of a computer, while 53.9% made use of the Internet from any location.

2002).[31] Growth of new users was slowing. If suppliers were going to experience growth, it would not be in “new users,” but in more intensive use by existing users.

As it turned out, a shake-out ensued, in which many small ISPs left the market or sold their assets to other vendors. Several large ISPs, such as PSINet went bankrupt from (what turned out to be) imprudent early expansions. Other ISPs (more commonly) merged with each other, perceiving that bigger firms had better chances for survival. Suhonen (2002) identifies 86 publicly known mergers in 1999, 152 in 2000 and 81 in 2001. There were many more than these.[32]

There is no evidence that geographic coverage declined, nor would that have been predicted for a technology with such small economies of scale in deployment. To be sure, the set of vendors decreased, in what many observers described as an inevitable and overdue consolidation of suppliers (Boardwatch, 2001).

It is hard to get a definitive sense of how many suppliers exited due to bankruptcies. Many sales and mergers were not publicized. Moreover, the industry for tracking the ISP industry underwent a shake-out along the same time as the industry, so there were no consistent reporting norms over time. Here are some indications of the type of change that occurred: In January, 1999, Boardwatch lists 4511 US ISPs. By 2001 that had fallen to approximately 3500 (Suhonen, 2002). Boardwatch lists just over 2500 for 2002. Most of this decline came from exit of small firms.

This decline in ISPs is plausible for another reason. After 2000 many other facets of the Internet experienced a decline in the rate of growth and a decline in the degree of optimism about future prospects. To be clear, this should not be interpreted as a decline in the level of demand for Internet service, but as a decline in the expected rate of growth of demand.

3.6 Variety and Quality Over Geography

As with many business ISPs pursued activities to develop more than just one source of revenue. Many ISPs provide services that compliment the physical connection. The most important and necessary service is an address for the user's computer. All Internet packet traffic has a 'from' and 'to' address that allows it to be routed to the right destination. An ISP assigns each connecting user with an address from its own pool of available addresses.

ISP’s offer other services in addition to the network addresses. These may include e-mail servers, newsgroup servers, portal content, online account management, customer service, technical support, Internet training, file space and storage, web-site hosting, web development and design. Many larger ISP’s also bundle Internet software with their subscriptions. This software is either private-labeled or provided directly by third parties. Some of it is provided as

[31] A survey of business establishments with over 100 employees showed that 88% of such establishments participated in the Internet in some form by the end of 2000. This clearly does not hold for small establishments, particularly those with less than twenty five employees and especially those with less than ten. See Buckley and Montes, 2002.

[32] For example, MainOne and Verio both were known for buying other ISPs in attempts to grow into a national ISP. Neither was commercially successful.

standard part of the ISP contract and some of it is not. Some ISP's also recommend and sell customer equipment they guarantee will be compatible with the ISP's access equipment.

It is obvious from industry publications that ISPs vary in the array of service they offer and the quality with which they provide them. How should this be measured? It is difficult to characterize the variance in either offerings or quality of a new industry. It is even more difficult to get information about the variance of provision in different geographic locations. The measurement challenges were particularly severe because official government reporting agencies had not standardized the classification of activity.[33]

A few studies provide insight into the interplay of geography and service quality. For example, Augereau and Greenstein (2001) look at the determinants for upgrade decisions for ISPs. They examine upgrades from dial-up–only service to 56K modem or ISDN service by 1997. In their model, they look for firm-specific factors and location-specific factors that affect firms' choices to offer more advanced Internet services. Their main finding is that "the ISPs with the highest propensity to upgrade are the firms with more capital equipment and the firms with propitious locations." The most expansive ISPs locate in urban areas. They further argue that this could lead to inequality in the quality of supply between ISPs in high-density and low-density areas.

Greenstein (2000a, 2000b) examines the variety of service offered by ISPs with different location coverage. His work focuses on the offering of additional services, such as hosting, networking, web design and high-speed access. Why do firms make different choices? Differences across firms arise when decision makers face different demand conditions, quality of local infrastructure, and labor markets for talent, or when they inherit firm assets of differing quality. These create a variety of economic incentives for adapting Internet infrastructure to new uses and applications.

Small ISPs in different parts of the country have very different propensities to offer additional services. Of the 1764 ISPs surveyed, 26.9% offer frontier access (i.e., they support provision of speeds higher than typical dial-up access at 56K), 22.9% offer networking services, 23.5% offer hosting services, and 38.6% offer web design services. Of the 325 ISPs primarily found in rural areas, 12.0% offer frontier access, 11.0% offer networking services, 13.8% offer hosting, and 23.3% offer web design services. The propensities for rural ISPs are between 40% and 60% lower in each category. Surveys of firms with national coverage (primarily in urban areas) find patterns similar to those small ISPs devoted to urban areas.

How should these patterns be interpreted? There appears to be significant ISP-specific factors that cluster together with the choice to offer new services. In particular, the size of the ISP, its geographic reach, some key capital

[33] Some of these difficulties were coincident with the US conversion to North American Industry Classifications in 1997. See E-Stats, as maintained by the US Census, at http://www.census.gov/eos/www/ebusiness614.htm. This activity begins tracking electronic commerce around 1999. The earliest official US data on the Internet is the official price index for Internet access in the US. It begins December 1997.

investments, and the ISP's focus on particular types of users and non-ISP lines of businesses all predict experimentation with new services. Greenstein (2000a) develops a statistical model that shows that the location-specific variables have less importance than the firm-specific variables except in the provision of high-speed access, where both factors are important. Interestingly, some of the firm-specific variables—whose distribution does vary over geographic space—also help explain some of the observed variance.

These findings are consistent with the view that small ISPs choose strategies for differentiation based on firm-specific growth strategies, limited (possibly) by indivisible assets arising at a local level. For example, in some cities, there are only a few locations for interconnection with an Internet backbone or a local telephone switch. Other equipment for running an ISP, such as a modem bank, could be cheaper at larger scale.

In other words, a small ISP in a rural area faces the limits of scale economies in facets of its operations. Scale economies arise from capital equipment investment and other fixed costs on the back end. Some times these were not severe when small ISPs rely on existing telecommunications infrastructure. Frontier access also can be scale intensive when it requires significant costs to set up. In that case, it may require higher volume of use to be profitable. Therefore, it may not be profitable without large capital investments and a density of potential users, typically businesses comprised of professionals, a situation prevalent in urban areas but apparent only in a few rural areas.

Ancillary services of the sort described above are also subject to some scale economies. Networking and web design require a core mass of business customers to defray the costs of acquiring capital and maintaining sufficient technical expertise. These costs are also defrayed by some economies of scope among these services and between these and other lines of business.

Overall, urban areas get more new services because of two factors: (1) increased exposure to national ISPs, who expand their services more often, and (2) the local firms in urban areas possess features that lead them to offer services with propensities similar to the national firms. That is, high-density areas almost always get some ISP entry, whereas some low-density areas get none or very little. High-density areas see an especially large amount of entry because they experience entry from nearly all the firms with national ambitions. In contrast, little or no entry in a low-density area virtually precludes availability of any complement to basic access. High-density areas benefit from repeated exposure to many ISPs that offer such services. More entrants will lead to more realized numbers of new services, thereby raising the probability of finding one, two or three instances of new services in a specific location.

Perhaps this is best illustrated in concrete terms. At the turn of the millennium small or isolated rural towns in the US were the only areas at risk for inadequate Internet service, if that had it at all. If they had some service, whether they got adequate service depended on small details about the local conditions. Did the local PC service guru want to run an ISP on the cheap in his back office or would he invest in a large modem bank? Would the local rural telephone company run Internet service if it barely broke even or operated at a loss? Would they only offer operational connectivity or would they also offer services associated with installation, designing web pages, and designing fire walls for local businesses? Did an Internet backbone connection happen to run along a nearby right-of-way, such as a major highway, pipeline or electrical line path? The denizens of some rural places got to experience the full benefits of the services affiliated with the commercial Internet because the right people were there and took the initiative to bring local services to the highest standards possible. In some places this did not happen the local population had to make do with less.

3.7 Summary

The experience of the dial-access industry offers one lens for filtering through the patterns of deployment of the Internet after commercialization. Over a decade there were many changes in the modal form of business. Changes in the delivery of services and changes in user expectations resulted in numerous qualitative changes in the basic service experienced by all users. The structure of the industry also fluctuated and a few dominant providers emerged in the second half of the decade.

In spite of fluctuation, a basic economic pattern emerged. Dial-up access was cheap to deploy because it built on top of the existing telecommunications infrastructure. It was a retrofit on the phone system, requiring a firm to organize access and a user to invest in modems. In virtually all locations except those with low density the costs of providing this service was minimal. So prices to households stayed comparatively low (around $20 on average). Closely related services grew on top of basic access, supporting the dispersion of Internet access to most potential users in the country. Within a couple years Internet service was nearly geographically ubiquitous.

4. The Location of Network Backbone

The commercial Internet is comprised of hubs, routers, high-speed switches, POPs, and high-speed high-capacity pipe that transmit data. These pipes and supporting equipment are sometimes called *backbone* for short. Backbone is comprised mostly of fiber-optic lines of various speeds and capacity. However, no vender can point to a specific piece of fiber and call it “backbone.” This label is a fiction, but a convenient one. Every major vendor has a network with lines that go from one point to another, but it is too much trouble to refer to it as “transmission capacity devoted primarily to carrying traffic from many sources to many sources.”

The Internet’s backbone connects servers operated by different ISPs. It connects city nodes. It transports data over long distances. The presence and structure of the backbone (potentially) provides information about which cities are playing the most prominent role in the development and diffusion of the Internet. The geography of these connections (potentially) provides insight into the economic determinants of the network. The connection to the backbone and the size of the backbone has an impact on local economic activity since it has an effect on a firm's ability to distribute large amounts of data via the Internet

What is the proper economic interpretation of the size of backbones going into cities? Comparing the backbones in different regions is not straightforward. Its presence depends on many things such as population size, type of local industry, and other facets of local demand. Its maximum flow rate only gets used at peak times and not most of the day, so statistics about capacity must be interpreted with care. Moreover, nobody would expect connection and bandwidth to be equally distributed across geographic space, so the appropriate benchmark for assessing this geographic dispersion of backbone was subject to debate.[34]

Despite this inherent interpretive ambiguity, there is intense policy interest in understanding the geography of commercially supplied backbone. There are concerns that some areas (e.g., small towns) were underserved while others (e.g., major cities) were served too much. There are also concerns that the industry was too concentrated, as well as nearly the opposite, that the competitive situation motivated firms to make impertinent and redundant investments. What is the relevant evidence and how should it be interpreted?

4.1 The Geographic Features of the Backbone

Prior to commercialization the backbone for U.S. networks grew out of the network built for the NSF. The NSFNet was designed to serve the research needs of universities. It connected over one thousand universities in North America by the early 1990s. It had concentrated much of its communications infrastructure around a dozen super-computer centers and other major research universities, which were distributed to as many different locations. That resulted

[34] For a variety of perspectives, see Kitchin (1998), Moss and Townsend (1999), Dodge and Kitchin (2001a, 2001b), Castells (2002), Malecki and Gorman (2001), Gorman and Malecki (2002) and an excellent review in Malecki (2002a).

in a geographically concentrated transmission network with a wide dispersion of access points. Frazer (1995) analyzes the location of NSFNet through its growth. Just prior to commercialization – i.e., in 1991 – the key end-points were university and research centers in Seattle, Palo Alto, San Diego, Salt Lake City, Boulder, Lincoln, Houston, Argonne laboratories (Chicago), Urbana-Champaign (central Illinois), Ann Arbor, Pittsburgh, Ithaca, Cambridge, Princeton, College Park, and Atlanta.[35]

The points of access, transmission and capacity concentration in the NSFNet were destined to be different from the concentration that developed after commercialization. The commercial Internet was not managed by a single entity. Under commercialization, in contrast, multiple suppliers made investments designed to support a myriad of targeted customer needs. What economic principles explain the geographic patterns of backbone capacity? What dimension of geographic location shaped the supply conditions facing users in different locations?

Motivated by the desire to examine the relationship between economic activity and Internet backbone, as well as cautiously aware of the uncertainty inherent in cybergeography, I argue that commercialization resulted in a network with two surprising features.

1. Resources for delivering the Internet concentrate in some areas so that there are overlapping networks serving the location. This redundancy leads observers to call the network "overbuilt." On close inspection, however, it is difficult to support this claim unambiguously.
2. The U.S. backbone network was built quickly and ahead of demand in some locations. In retrospect the speed with which it was built appears almost reckless to some observers. On close inspection, however, there was an economic rationale to much of this behavior, so, once again, it is difficult to argue that behavior was unambiguously "reckless".

4.2 The Industrial Organization of the Commercial Backbone

The geographic features of the US backbone need to be understood in terms of the industrial organization of interconnection in the commercial era. Several factors are salient.

First, the commercial Internet developed distinct cooperative institutions for exchanging data. To be sure, this observation has to be qualified for the time period and the activity. As the Internet made its transition into private hands, cooperative principles persisted due to a combination of good policy and – for lack of a better phrase – corporate ethics. An example of good policy: The commercial network retained several of the public exchange points developed by the NSF, and transferred their operation to associations of private firms.[36] An example of ethics: descendents of the NSFNet were managing the networks at

[35] See http://archive.ncsa.uiuc.edu/SCMS/DigLib/text/technology/Visualization-Study-NSFNET-Cox.html or Frazer (1995, especially p. 33).

[36] For extensive discussion, see Mackie-Mason (1995), Srinagesh (1997), Bailey and McKnight (1997), Chinoy and Salo (1997), Cawley (1997), Farnon and Huddle (1997), National Research Council (2001). Also see Milgrom, Mitchell and Srinagesh, as well as Besen, Spigel, and Srinagesh (2001) for a more recent update to this line of thinking.

MCI, UUNet, BBN, or IBM and elsewhere. None of them tried to exploit their strong leading positions with strategic denial (or overt degradation) of interconnection to new entrants. Such behavior could have affected private entrants with aggressive ambitions, such as PSI (whose management was also descendent) or at a later time, Level3, who built an entirely new national backbone. That is, incumbent firms did not refuse to interconnect entrants as part of a discriminatory policy designed to foreclose entry.[37]

Second, the tenor and tone of cooperative behavior started to change in the mid 1990s as the commercial prospects for the Internet became more apparent and demonstrably larger. For example, a few large firms exchanged traffic but did not charge each other for it, nor discriminate on any basis other than size of data traffic. At the same time they announced policies to charge for interconnection. The practice of not charging became known as private peering.

There were a number of facets to these arrangements. Peering meant that firms exchanged data without the hassles of tracking traffic in order to come up with explicit compensation. Several of the largest data carriers exchanged data this way if the volumes were within (depending on the firm) two to eight orders of magnitude of each other. MCI/UUNet was among the largest backbone providers to implement this policy in the mid 1990s. It justified this policy on the basis of their engineering efficiencies.[38] That is, monitoring traffic levels at high volumes was costly. Peering agreements freed both parties from monitoring each others' traffic.[39]

While at first the seemingly genteel nature of peering was regarded as a legacy institution from the pre-commercial era, more concerns arose when the largest firms started peering with only large firms and charging smaller ISPs for access, using contracts of varying lengths and obligations . This meant that the large firms could *de facto* price discriminate, charging (effectively) higher prices to small firms and lower prices to the firms with whom they peered. Given the potential costs savings of peering, such price discrimination was not necessarily anti-competitive, but it did start to raise questions among regulators about large firms with market power treating rivals differently.

Moreover, smaller ISPs had to exchange traffic at the public exchange points. This raised the further question about whether public exchange points were as fast as private peering (if these were equal, why did the large firms prefer to peer away from the public exchange?). It also raised issues about whether the users of smaller ISPs received lower quality service because their

[37] That is not say there were no conflicts of the public and private. See Kende (2000) for a discussion about the entry of Level 3 and their complaints about the peering policies of UUNet. Kende concludes that much of this complaint arose from Level 3's dislike for purchasing transit services from UUNet. UUNet refused to peer at no cost until Level 3's actual volumes reached their ambitions.

[38] The engineering justification for peering was seemingly straightforward. As long as volumes of traffic exchange were close to each other in magnitude, there was little business sense for either party to closely track volumes of traffic and charge for transit services.

[39] Coffman and Odlyko (2002), page 20, assert that most traffic by 2002 travels through private peering, in sharp contrast to seven years earlier, when most traffic traveled through public exchanges.

traffic exchanged at slower rates.

Because the technology was new and because NSF had commercialized the Internet with only general guidelines about Internet connection practices, there was not a clear regulatory policy for thinking about these issues. Yet, in the background was the vague memory of AT&T self-serving interconnection policies almost a century earlier, feeding a larger public policy concern about the ease with which new firms could enter and interconnect.

While there was no overt evidence of such behavior by the largest backbone providers, MCI or later WorldCom, there were always concerns about abuse of market power. Regulators were sensitive to concerns that market leaders would not invest aggressively to serve new areas or show attention to small firms requesting interconnection in difficulty circumstances.[40] Such concerns motivated one of the conditions for the merger between MCI and WorldCom, that the merged firm sell some backbone assets.[41] Similar concerns motivated intervention to prevent the proposed WorldCom merger with Sprint.

Yet, this reasoning necessarily oversimplifies the implications for the geographic reach of the Internet. Some elements of interconnection pricing – between two ISPs, for example – were entirely market oriented and mediated by contracts, while other facets were heavily regulated – such as between and ISP and an incumbent local exchange company (ILEC), especially after the 1996 Telecommunications Act. In effect, no participant in this market ever really escaped the effects of this regulation. Some transactions were simply more distant than others. Would efficiency in one part of the system be in society's interest if many other parts of the network were priced according to non-economic principles? Or would it simply mean that a few large firms (and their predominantly urban users) were benefiting form the distortions imposed on others?[42]

Once the commercial Internet began to explode in size and number of participants with new entry after 1995, a commercial intermediary became prominent and changed the relationship between geographic location and interconnection. Popular web sites, such as Yahoo or AOL, and retailers who saw competitive value in speed, such as Amazon, made deals with cache/mirror sites operated by firms (e.g., Akamai, Digital Island, among others). Different intermediaries operated in different ways, but all had the same purpose – eliminating performance degradation from national backbone congestion and data handoffs at switching points. Congestion on the national backbone could degrade quality if it slowed the movement of packets from a host site in one place, say Seattle, to a distant place, say New York. Caching eliminated much of the differences in performance between locations in the US—at least for the most popular sites and ISPs and content providers who could afford it (Carr, 2000).[43]

[40] For theoretical underpinnings see the discussion in Laffont and Tirole (2000), or Laffont, Marcus, Rey andTirole (2003).

[41] Eventually MCI's Internet backbone was sold to Cable and Wireless.

[42] For discussion, see, e.g., Noam (2001), Noam (2002), Laffont and Tirole (2000), Kende (2000), and Sidak and Spulber (1998).

[43] For example, Akamai's network maps the entire Internet every few minutes, identifying points of congestion. If one server is down, data requests are handled by other servers. Then Akamai's

Users in New York down loaded popular material nearly as fast as users down loaded material in Seattle. Users in all but the most isolated small towns could have as fast a service for mainstream applications as users in central cities.

This intermediary's presence was quite important for the location of backbone. These deals eliminated much of the competitive differences between ISPs in different locations and the backbones supporting them. The difference between a local and national ISP came down to software hosted on the ISPs' server and embodied in its connection with other ISPs. To be sure, it only held for the most popular sites and the largest ISPs, but that was still a high fraction of the experience of most users.[44]

Finally, one other institution deserves attention: indefeasible rights of use (IRU). Firms can obtain either space in conduits or dark fiber (i.e., fiber *sans* terminal equipment), typically for twenty years at a time. This permits a firm to buy the option to have distribution capacity in a certain location, deferring investments in equipment, switches and other assets affiliated with operations. IRUs became common in the late 1990s, as firms with rights-of-ways in one location leased IRUs to firms in another, and visa-versa.

IRUs were important for three reasons. First, it limited the sunk investments a firm had to make in order to secure the option to providing service in an area when their customer base was small in the short run. This was a cost-minimizing strategy in an era of potential growth and demand volatility (Faulhaber and Hogendorn, 2000) Second, it fostered the impression in policy discussion and popular discussion that the US backbone network was more redundant than truly was the case. That is, the independent statements from firms about their network footprint could lead to errors of double counting if IRUs were not taken into account. For example, Hogendorn (2003) shows that such IRUs accounted for more than half the new *growth* in new fiber between 1997 and 2001, though this was commonly not stated in news accounts in the *Wall Street Journal*, among other publications.[45] Third, these types of contracts eased potential entry by the largest firms into different territories, heightening competitive pressures between them. Hogendorn (2003) speculates that this behavior also contributed to the rapid decline in prices for backbone services.

4.3 Interpreting Networking Practices

The non-commercial Internet had begat a commercial Internet that retained some features, such as peering and end-to-end, and adopted new features, such as commercial caching. How were firm practices to be understood

servers send the user all the low-bandwidth elements of a page, such as text. It also instructs the user's browser to get high-bandwidth content, such as photographs, from geographically close servers. Akamai's algorithms determine the optimal servers and routes for each end user.

[44] Web site surfing among non-AOL users during the early period of commercialization, as now, was remarkably concentrated. Several hundred sites accounted for the vast majority of the time on line. Use of portals alone accounted for more than a quarter of the time on line among non-AOL users (Goldfarb, 2004). Since AOL use was at least 40 percent of household use of access, a large fraction of user experience interacted with a mirror/cache site in some form.

[45] He also points out that the FCC discontinued collecting data about the state of fiber networks in the US after 1998, so there was no government oversight to correct this common misperception.

and interpreted? Such questions went to the differences in behavior between the public/private domains in the commercial Internet – e.g., whether the privately managed Internet required public intervention or not because it did or did not accomplish public goals, such as geographic ubiquity.

To the delight of some and the dismay of others, these factors supported a backbone that some called *tiered* (Frieden 2001, Kahin, 1997). The tiers were associated with size of footprint and volume of traffic. Tiers became a short-handed designation for which firm carried data over long distances and which collected charges from others for transit service. For example, the largest national firms all became tier-1 providers. This included AT&T, IBM (before being sold to AT&T), MCI (whose backbone was sold to Cable and Wireless), UUNet (eventually sold to WorldCom), Sprint, and Genuity, among others.[46] Most regional and local ISPs became lower tier ISPs, purchasing interconnection from one of several national providers.

Not all analysts argued that tiers had a hierarchical interpretation. Others observers – who were skeptical of the rigidity of tiers – called this system a *mesh* (Besen et al, 2001). In this view, participants faced many options for interconnection. For example, many ISPs arranged to use multiple backbone providers (i.e., known as multi-homing), significantly diminishing the market power of any particular backbone provider in any given location. Related, dial-up ISPs with large networks (e.g., such as Genuity) could receive calls in one location but back-haul it to another where it is then connected to the Internet. This type of network design gave ISPs multiple options for connecting to the Internet, limited the discriminatory power of any single backbone firm.

Whatever it was called, this system was not associated with overt discriminatory treatment in quality of service across location. That is, the system had remarkably little effect on the quality of applications in different locations, at least at first. The engineers called this the preservation of the "end-to-end" features of architecture (Blumenthal and Clark, 2001, National Research Council, 2001). Performance was not based on either the identity of the end user, the user's location, or the type of application. Stated simply, transmission capacity built anywhere interacted with transmission anywhere else without altering the application.

The preservation of "end-to-end" partly reflected economies of scale from standardization. Contracts for different bandwidth and priority handled the costs of interconnection between ISPs and backbone firms, but these transactions did not greatly alter the user's experience of the application.[47] Ownership of facilities did not induce discrimination on the basis of the origin or destination of the data or type of application.[48] It was not in any carrier's interest to use non-standard

[46] Among the others sometimes counted as tier-1 providers include Qwest, IXC, Williams, and Level3. See Kende (2000) or Hogendorn (2003) for different discussions.

[47] This is somewhat of a simplification. ISPs who contracted for larger capacity and first priority did offer fewer peak-load issues to downstream users. For many uses, such as e-mail, these contracting practices had little impact on the user experience.

[48] See Kende for a careful description of how interconnection takes place. See Besen, Spigel, and Srinagesh (2000), and Besen, Spigel, and Srinagesh (2001), for an argument that the lack of discrimination in interconnection resulted from a combination of peering, multi-homing (ISPs

equipment or operations, nor to vary from the common protocols, lest they lose access to the possibility of imitating innovative practices arising elsewhere.

To be sure, many Cassandra-like observers foresaw several different types of threats to the preservation of this end-to-end feature from several different types of applications and proprietary solutions (and still do see such threats on the horizon).[49] And, yet, widespread incompatible balkanization or discriminatory interoperability had not yet emerged in the first decade after commercialization.[50]

Rather, the presence of end-to-end accompanied several behaviors that shaped the Internet backbone's geographic layout. First, vendors were not discouraged from building geographically overlapping networks, since such action did not alter whether they ultimately interconnected. Related, it also encouraged the use of IRUs for covering new geographic territories, since using such assets was standardized across geographic regions.

Second, vendors also could build the opposite of a national footprint, namely, specialize with regional footprints but still ultimately interconnect. Or, related, a regional firm could choose to build its own facilities in a region, but contract for backbone capacity in other regions in the event that their clientele valued wider geographic coverage. Again, interconnection was assured, so a national footprint was not necessary for survival.

Third, the firms with national footprints did not have big advantages in the deployment of end-to-end applications that required a low amount of signal delay, such as virtual private networks, video conferencing, or applications requiring tight security. In other words, there were few highly valued transactions that were more efficiently done under the ownership of a single firm with wide geographic reach.

In closing this discussion about practices, it is important to note that these patterns touch two open questions at the core of network economics. Many economic models of non-interconnecting networks suggest that firms would build geographically redundant facilities in order to compete for customers. Yet, the Internet networks of the late nineties interconnected as well as any observer could imagine, and redundancy still emerged in many locations. In this case, such redundancy seems to have emerged as a result of competition for customers (more on this below) and deliberate government policies to prevent mergers resulting in concentration of ownership. It would seem that redundancy is inevitable under any structure for competition, whether or not networks interconnect. Is there a grain of economic logic to this observation?

Second, economic arguments about natural monopolies highlight the need

---

using more than one backbone provider) and other behaviors generating multiple alternatives for users. In their view, these features eroded market power of all backbone players, fostering incentives for non-discriminatory interconnection.

[49] These arguments became especially heated during the regulatory hearings for the AOL-Time Warner merger. Proponents and opponents of "open access" foresaw either grave threats or little concern on the near horizon.

[50] See the discussion in Blumenthal and Clark (2001), Lemley and Lessig (2001), as well as in Kruse, Yurcik, and Lessig (2001) on why end-to-end may end in the next generation of applications.

to eliminate redundant lines in order to have low-cost operations. Indeed, it is a key argument for establishing a single provider of a natural monopoly. Yet, the Internet of the late 1990s was as far from a natural monopoly as any observer could imagine. Entry was rampant, so was growth, and so was redundancy in geographic breadth. Did the dynamic benefits from this entry and growth outweigh the potential costs of such redundancy?

The following sections illustrate these open questions in a variety of dimensions.

4.4 Interpreting the Geographic Dispersion of Capacity

During the first few years of commercialization of the Internet a handful of cities in the United States contained the majority of backbone capacity (Gorman and Malecki 2000, O'Kelly and Grubesic, 2002). Specifically, San Francisco/Silicon Valley, Washington, DC, Chicago, New York, Dallas, Los Angeles, and Atlanta contained links to the vast majority of backbone capacity. Depending on how it was measured, sometimes the relative ranking varied, but the list of the top seven does not. As of 1997, these seven cities accounted for 64.6% of total capacity. By 1999, even though network capacity quintupled over the previous two years, the top seven still accounted for 58.8% of total capacity.[51]

From the outset, distribution of backbone capacity did not perfectly mimic population distribution within metropolitan regions. Seattle, Denver, Austin, and Boston have a disproportionately large number of connections (relative to their populations), whereas larger cities such as Philadelphia and Detroit had disproportionately fewer connections (Townsend 2001a, 2001b). In addition, the largest metropolitan areas are well served by the backbone, while areas such as the rural South have comparatively few connections (Warf 2001).[52]

Cities such as Boston and Seattle also experienced favorable outcomes in terms of growth. Grubesic and O'Kelly (2002) measure the speed of growth at metropolitan areas. Their data indicates that areas such as Milwaukee, Tucson, Nashville, and Portland had large growth in POPs at the end of the 1990s.

Despite the growth of the total capacity, maps of Internet backbone for cross-national data transmission have not changed much from 1997 to the 2002. Maps of Internet backbone at both early and later periods display similar geographic features in the biggest arteries. A few key cities have the greatest number of connections. The main difference arises in the minor capillaries and passageways of the network, which are more abundant later. [53] To visualize this idea, a few representative Maps from AT&T and UUNet from 2000 are included in Figures 4 and 5.

---

[51] Gorman and Malecki (2000) look at each of the ten major backbone networks. They compare them using measures of median download time and the number of routes available to an ISP. Data was provided by the 1998 *Boardwatch Magazine's Directory of Internet Service Providers.* Malecki (2002) provides a summary of this and related research. They do not correct for the double counting identified by Hagendorn (2004).

[52] Employing graph theoretical analysis, Wheeler and O'Kelly (1999) rank metropolitan statistical areas (MSAs). Their results are qualitatively similar to the ones mentioned above.

[53] See also the maps of the US network at www.telegeography.com. See Dodge and Kitchin (2001a, 2001b), or www.cybergeography.org./atlas/atlas.html.

At first glance, economic reasoning offers a variety of straightforward explanations for these patterns. For example, there are economies of scale in high-capacity switching and transmission, even during a period of high growth. There are also economies of scale in data-interchange. [54] In other words, as with many communication networks, the present backbone network is a hub and feeder system with a few hubs.[55] These hubs retain classic economic features of all traffic hubs: the variable costs of operating the service are lower when their transit depots and switching functions are close to each other.

Classic hub economics may explain why hubs will arise, but it does not explain where they settle in one place and not another. For that latter question, three factors seems particular salient for the Internet. For one, some of the land-line telephony firms were also in the business of carrying data traffic (e.g., AT&T, MCI, Sprint). Their voice and data network configurations would naturally use some of the same equipment, facilities, etc, so the locations of their high-speed lines would not change much, if at all.

Next, whether operated by an incumbent, or by a new entrant such as Level3, there are only a few (commonly employed) right-of-ways pathways available for long-distance transmission lines. These are rail lines, highways or pre-existing pipelines or electrical lines. Said simply, there are only a few ways to get across the country, so major trunk lines for each commercial network follows similar paths.

Related, similar right-of-way considerations shape the deployment of high-speed lines within any regional area or municipality. Public transit lines, freeways, pre-existing underground tunnels, and pre-existing telephone/electrical poles and structures act to limit the number of pathways to any given location from any given location. Access to these right-of-ways is available on a non-discriminatory basis through a licensing process operated in most states.[56] So, again, it should come as no surprise that the same firms employ the same pathway and interconnect to regional networks in a few places.

Many of the choices for the location of NAPs – made by NSF or the earliest entrants – have persisted into the commercial era either as a public exchange point or as a focal location for clusters of private peering. To be sure, some of these points, broadly speaking, would have arisen under any system because their location was determined by the centrality of some places in the midst of traffic flow. For example, data-exchange somewhere near New York or Washington DC makes sense for traffic along the east coast. One can make similar remarks for Dallas or Atlanta in the south, Chicago in the Midwest or San Francisco and Los Angeles in the West.

[54] There is an additional economy of scale at public interchange points in the sense that all parties benefit if more parties partake in interchange at the same location (see Srinagesh, 1997 or Kende, 2000).
[55] Moss and Townsend (1997) observe that the earliest commercial backbone for the Internet is organized as "hub and feeder system with distinct nodes, contrary to popular notions of a dispersed, chaotic Internet."
[56] There is also federal legislation regulating attachment to telephone poles. This legislation has a long history going back to the regulation of Interconnection with AT&T and the diffusion of cable television firms. The 1996 Telecom Act also addressed some of these features.

One might see related factors shaping the service into some cities. Proximity to major telecommunication centers explains the findings in Tucson and Milwaukee (i.e., near Los Angeles and Chicago, respectively). As another example, some cities, such as Portland, Ore., are located between larger cities with high Internet activity (i.e., between San Francisco/San Jose and Seattle). Some cities are simply blessed with centralized locations between many points of traffic, such as Nashville, and, on a larger scale, Chicago.

With a classic transportation hub, the city that acts as the home gets large benefits in the form of jobs, infrastructure, tax revenue, and local support services. Did the location of Internet hubs attract complementary economic activity that led to economics benefits for the local area? In practice, the question is impossible to answer on the basis of what happened in the US in the first decade after commercialization. Some areas that naturally attracted hubs – such as Washington D.C., New York, or San Jose – would have had large network markets anyway. There was high demand and large supply naturally followed.

At the same time it was not at all obvious that a vendor supporting a web page in Akron Ohio got markedly worse service than one in Los Angeles or San Francisco because this depended on matching supply and demand. The key issues concern whether larger network capacity in a backbone (using high capacity OC speeds) makes any difference to a host site or to a user, who ultimately connects at much lower speeds. Since the majority of users connect at much slower speeds than the highest speed backbone, the weakest link may have mitigated any advantage from geographic proximity.

As it turned out, the dot-com crash in the spring of 2000 interrupted the boom in complementary activities, such as the growth of data-warehouses next to fast backbone near Chicago. Moreover, the boom and bust happened so quickly, it also reshuffled the labor market for Internet services and consulting – both up and down. It is also unclear what effect will result from the bankruptcies of some of the firms who provided backbone services. These open questions will only be answered with more time.

Finally, who gets service at all? Since the Internet was experiencing explosive growth it is highly doubtful that any city over a quarter million was not served by several backbone firms. For small cities the relevant issues were similar to that which determined ISP coverage, such as density of demand or proximity to major conduits along highways.

4.5 The Economic Interpretation of Redundancy

Relative to its history and origins, the appearance of redundancy in the backbone was notable. For the Internet, different firms replicated transmission capacity along similar paths to the same cities, thus they potentially serve the same customer. This was not a feature found in the NSFnet.[57]

But history probably offers a misleading point of comparison for understanding economic geography with commercial suppliers. The more

[57] The motives for redundancy are quite different in a military network communications network, where the communications lines must survive potential damage to parts of the whole. Only a little research has examined redundancy in this sense. See Grubesic, O'Kelly and Murray (2003).

relevant question is whether this outcome is remarkable in comparison to other industries. In many industries observers *would not* find it surprising that capacity from different firms serves the same customer. Commercial firms in most industries do not coordinate their building plans with each other and, consequently, overlap in their aims. In many industries different firms provide similar products to similar sets of customers. It is important to recognize that some features of the redundancy of the backbone arose from competitive forces introduced into the Internet after 1995, which are otherwise normal for many other markets.[58]

Several factors were acting simultaneously. They pushed in the same direction and enhanced the incentives to grow quickly: (1) competitive incentives; (2) impatient investment behavior on top of normal competitive behavior; (3) optimistic assessments of future demand growth and (4) beliefs about the strategic advantages of first-mover actions. I address each of these in turn.

For one, the backbone network of the late 1990s embodied features that reflected strongly on the presence of competitive incentives. The U.S. backbone was built in increments by many firms, not by a single firm with monopoly ownership. There was a multiplicity of actors building it and competing with each other. In the late 1990s, notably AT&T, Sprint, WorldCom, GTE, Level3, Qwest, Global Crossing, Cable and Wireless, and Williams, among others, all had well-known plans to build networks with national geographic footprints.[59]

Competition enhanced the incentives of each actor to grow quickly, price competitively and experiment broadly. As in many markets competitive incentives induced suppliers to identify and tailor network services to unfilled customer needs. In a few locations this competition increased the competitive options available to potential users and ISPs. Thus, it was quite common for ISPs to seek and sign multiple agreements with backbone partners (Boardwatch, various years).

There is no question that the impatient financial environment of the late 1990s provided further strong incentives to grow quickly. Part of this was due to Wall Street's exuberance, as stock prices responded (sometimes without sufficient skepticism) to announcements of building plans, new Internet ventures that generated backbone investments, even when it resulted in significant redundancies between rival suppliers. Certainly this reinforced the previously noted incentives to *announce* plans for expansion, even when it just involved signing an IRU.

And, yet, part of this impatience was real. Businesses did invest in TCP/IP applications. The latter half of the nineties witnessed the single greatest growth in investment in Information Technology in the history of the US. Investment grew at 20% a year from 1995 to 2000, with almost three quarters of that being business investment.[60]

---

[58] For an interesting comparison of the differences between factors leading to boom and bust with Internet backbone firms and the classic railroad boom and bust, see Hogendorn (2003).
[59] For a sense of these plans, see Boardwatch directories, various years.
[60] This includes computer hardware, computer software, communications hardware and instruments. By 2000 computer hardware and software stocks had reached 622.2 billion (1996) dollars. See Price and Mckittrick (2002) or Henry and Dalton (2002).

In the face of this growth there was a common perception among suppliers that demand for Internet traffic would grow quickly for many years. Again, it is hard distinguish between the reality, hype and dream that underpinned this demand. The late 1990s gave rise to a pervasive technological deterministic rhetoric in business media outlets. Decision-makers such as senior executives didn't exactly know what that future was going to look like, but they all believed that technology was going to play a central role in it. So they pursued an array of technology-related initiatives hoping that one or more of these would contribute to company growth or even mere survival. It all translated into extraordinarily loose budgets for information technology projects in business.

Forecasting had been quite sober when the Internet first commercialized (e.g., Meerker and Dupuy, 1996), but the restraint was lost in the late 1990s. In some circles, forecasts included seemingly unrealistic growth forecasts – almost without bounds – and in retrospect one has to wonder who could have trusted them enough to invest serious money on such flimsy forecasts. In some circles, consultants backed up projections with statistical samples. For the latter, it is significant that, well prior to its bankruptcy, WorldCom's broadband division had claimed that its Internet traffic was doubling every six months, a canard that became repeated often. While there may have been some truth to this report at one time early in the explosive growth of the commercial Internet, there was little confirmatory evidence of such persistent growth. Yet, this phrase was repeated often well up until the crash of 2000.[61] In other words, traffic grew, but many economic actors—such as Wall Street analysts, financial investors, and large firms actually building the networks—believed it would grow even faster than it did.

Finally, there was a common understanding that savvy business practices required laying backbone capacity (i.e., sinking fiber-optics into the ground) ahead of buyers' demand for that capacity. Firms perceived that business users of capacity would be hesitant to switch suppliers once contracts were signed. They perceived that small ISPs would not want to renegotiate new contracts with new suppliers. Backbone suppliers saw themselves racing with each other to sign (what at the time was forecast would be) lucrative contracts for services with low variable operating expenses. Hence, all vendors perceived themselves to be in a race to build volume, that is, sign-up users today, who would later foster profitable revenue streams.

This behavior was more consistent with the aggressive practices of venture funded firms, and it was far from the cautious building practices found in the regulated parts of the communication industry. In a regulated industry caution is a virtue, particularly if investment decisions can be second-guessed by

[61] In retrospect, it appears that this observation had some truth to it in 1997 and 1998, when the Internet first commercialized into a mass market, growing from a small base. While there certainly was growth in data traffic after that, it is unclear when growth rates began to *slow down*. It is also unclear whether WorldCom or others repeatedly asserted that these observations grew as fast as they asserted due to oversight or, deliberately in order to foster their own business plans. In the heady days of Wall Street boosterism for all things associated with Internet, such assertions were encouraged, and often went unquestioned and untested. For a more skeptical view, see the discussion in Odlyzko (2001), or Sidek (2003).

regulators. Capacity is built only after it was requested by buyers and approved by regulators. The difference can be stated simply: When a firm is competing for customers, less cautious investment practices are called for. To be sure, a stopping rule for such aggressive investment is hard to articulate. As in any industry with high sunk costs and (resulting) low variable costs, such aggressive investment is quite risky for any individual firm.

It should be no surprise that this incautious competition between network providers fostered uncoordinated build-outs among rival firms, which, in turn, fostered overlapping footprints and other redundancies in supply at a broad level. This was particularly so when firms signed an IRU, i.e., when they just rented the option to enter without building the entire capacity to operate (Hogendorn, 2003). Not surprisingly, commercial firms did not coordinate their building plans with each other, and they replicated potential capacity along similar paths. Said more concretely, every firm with a national footprint thought they had to be in San Francisco, Chicago, Los Angeles, Washington, DC, New York, and many other major cities. Stated simply, the rapid and redundant build-out of the late 1990s arose from one and the same set of competitive incentives.

4.6. Boom Leads to Bust in the Backbone

A decline in optimism did eventually arrive. This was partly the result of saturation. By 2000 most of the strong "first-adopters" among businesses and households had adopted the Internet.[62] If further growth was on the horizon, it had to occur through capital deepening by existing users, not through addition of new users. Indeed, business investment in IT reached its peak in 2000 and dropped after that.[63] New household of the Internet also slowed at the same time.[64]

The trade press dates the beginning of the decline of optimism at the spring of 2000, when financial support for dot.coms collapsed. Pessimism reached a nadir in the fall of 2001, after the September 11 terrorist attacks shook business confidence in long-term investments. This low plateau continued as the WorldCom financial scandal became publicized in the spring of 2002. Consequently, some backbone providers curtailed the expansion plans they announced in 1999 and previous years. Others left the market altogether.

As of this writing, this down cycle is not over, which leaves an open question about the long term shape of the U.S. backbone networks. Qwest, Level3, Sprint, Global-Crossing, MCI-WorldCom, Williams, PSINet, AT&T and others all invested heavily in redundant transmission capacity during the boom. Most of them have not fared well. A few, such as PSI and Global Crossing, have

[62] The results in National Telecommunications Information Administration (2000, 2002) indicate that by 2000 most households who had a computer had considered adopting the Internet. Most new growth in adoption was coming from households who needed to buy a computer. See also Goolsbee and Klenow (1999). Forman, Goldfarb and Greenstein (2002) find that most business establishments had were participating on the Internet by the end of 2000.

[63] See Price and Mckittrick (2002) or Henry and Dalton (2002).

[64] See National Telecommunications Information Administration (2001). Capital deepening at that point took the form of cannibalization of dial-up with broadband. See the further discussion below.

exited, selling their assets to others.

Signs of financial distress are evident in many places. Here are some examples: (1) Much of that fiber went unlit after its installation (as one would expect if expectations were over optimistic). (2) UUNET, a division of WorldCom, was the largest backbone data carriers in the United States, but the scandals at WorldCom led to its bankruptcy and cessation of investment activity (though the stress did not necessarily signal problems with the Internet division). (3) AT&T bought TCI and MediaOne, invested heavily in upgrading cable networks for purposes of offering household telephone and broadband access, but did not get the type of demand expected. Their CEO had articulated a grand design for the future and he was forced to resign when the strategy clearly failed. The strategy was repudiated and the cable assets were sold to ComCast. (4) PSINet, another early pioneer in building the commercial Internet backbone, overextended itself and had to declare bankruptcy. Its assets were sold to other investors at a severe discount. (5) Severe stress was also felt upstream at switch and equipment makers, such as JDUniphase, Corning, Cisco, Nortel and Lucent.

While nobody expects consolidation of ownership to completely eliminate redundancies, this downturn will reshape the ownership overflows of data traffic in the future. Will consolidation eventually lead to the emergence of this networks' dominant executive, as Cornell was for the telegraph and Vail was for the telephone? The final configuration of ownership remains in flux, awaiting the results of sales of assets for bankrupt firms, as well as, perhaps, the vision of temperate executives.

4.7 Interpreting a Decade of Building

Would the U.S. backbone network have experienced different geographic coverage with a different mode for organizing commercialization? The answer depends on whether one looks at the network before or after it is built. In other words, at the outset there was a major concern about whether the network would ever be built at all but little concern for whether it would be oversupplied in one place or another. After the network was built, overlapping footprints and other redundancies appear inefficient, though they were of little concern prior to its building.

A competitive market gives every actor strong incentives to build its networks quickly, price it low, and customize it to user needs. The backbone might not have been built as rapidly in the absence of such competition. Similarly, uncoordinated investment of sunk investments has the potential to lead to price wars in the event of overbuilding. But price wars cannot arise unless firms build their networks in the first place.

It would be surprising if uncoordinated investment resulted in too little infrastructure during a period of sustained demand growth, as had occurred in the late 1990s. Building ahead of demand is a calculated gamble for each provider. Every actor risks winning the same set of future customers. If all actors had an optimistic assessment of the future, then when that optimism declines, all will be caught with excessive assets. Of course, these incentives are further exacerbated if growth forecasts (from rivals or industry analysts) exaggerate true trends, as the forecasts coming from Worldcom did during the late 1990s (Odlyzko, 2001).

There was a sense of inevitability to the price wars that eventually erupted – i.e., the oversupply of capacity compared with demand. High initial demand in a market with uncoordinated investment should lead to a period of intense growth followed by a period of regret. The period of regret is almost inevitable, particularly when the growth is sustained by unbounded optimism and fueled by incautious building of sunk assets used in a race to search for customers.

The concentration of infrastructure in a few cities, while partly the result of classic hub economics, also looks like the natural outcome of the same speculative process during a period of speculation. All firms are rivals, investing in the same corridors for carrying data traffic between major population centers. Each wants to assure that its traffic receives priority from its own operations, not that of its close rivals. [65]

4.8 Summary

I conclude that the concerns about excessive market power on the backbone did not affect the growth and diffusion of Internet infrastructure during the first decade after commercialization. This might have been an issue, and there are many reasons why it did not manifest. The most important was that no firm – notably, not even MCI/WorldCom – amassed enough market share to dictate terms in most places.[66] In most major cities ISPs had many options for sending traffic to the Internet. In sum, I conclude that interconnection issues have not shaped the geographic dispersion of the Internet in the U.S. – at least, not yet.

The speculative investment behavior of backbone providers underlay an extraordinarily rapid build-out, an outcome that clearly benefited society. This same behavior laid the foundations for a financial bust and restructuring. There is (as yet) little persuasive evidence that it has led firms to massively substitute away from investing in peripheral locations. In addition, there has not been a domino effect of one bankruptcy causing another, even with a bankruptcy from WorldCom, one of the largest backbone providers in the world. In this sense, I also conclude that this boom and bust cycle has not been a central policy issue for the geographic dispersion of the Internet in the U.S. – at least, not yet.

[65] Similarly, it also suggests that the infrastructure from every major provider serving central locations, such as Washington, DC, should be measured according to benchmarks distinct from those used to measure infrastructure from several providers serving a less central location with much lower aggregate demand.

[66] Clearly this outcome was due, in part, to forced divestiture. As a condition for the MCI and WorldCom merger, MCIs Internet assets were sold to Cable and Wireless. The WorldCom and Sprint merger was (effectively) blocked by the European Commission over these and related issues (before the US Department of Justice registered its stance).

5. The Growth of Broadband

Internet access is a geographically local and non-tradable service. As a dial-up service, it is cheaper if it is part of a local phone call. If it is a broadband service, it is simply not available unless a local supplier has invested in the necessary infrastructure to bring delivery to a home or business[67]. In contrast to dial-up concerns about backbone, the concern for broadband centers on the fast speed, high price and limited coverage of available service.

By the late 1990s there were three basic delivery modes for broadband. The first was direct supply of high-speed lines, such as T-1 lines. This was prohibitively expensive for all users except businesses, and even then, it was mostly used by businesses in dense urban areas, where the fiber was cheaper to lay. Two other options became available in the latter part of the decade after commercialization: cable Internet or DSL.[68]

Three questions shaped analysis on where either provider would make investments in Internet infrastructure:

1. Cable and DSL diffused disproportionately to urban areas initially. What factors encouraged this outcome and was there any indication that this would change?
2. Did broadband have features that would eventually lead it to diffuse with the same speed and geographic ubiquity as found in dial-up Internet access?
3. Did the growth of new regulatory rules under the 1996 Telecom Act alter the geographic diffusion of broadband?

5.1 Why Broadband Favors Urban Areas

Broadband access is defined by the FCC as "the capability of supporting at least 200 Kbps in the consumer's connection to the network," both upstream and downstream (National Research Council, 2002, Grubesic and Murray, 2002). Whereas dial-up connection has moved past the frontier stage and is approaching geographic ubiquity in the United States, broadband access is far from ubiquitously available or adopted. It is much more common to large businesses. Its ubiquity will require quite a build-out because it is a local good in the minute sense—at the level of the block within a city. Sometimes this is called the "last mile" problem.[69]

The appeal to users of broadband is well known. It is faster than dial-up

[67] Advanced telecommunications services for packet switching involves a multiplicity of technologies, such as switching using frame relay or Asynchronous Transfer Mode, as well as Synchronous Optical Network equipment or Optical Carrier services of various numerical levels. Their deployment is not the central focus of this discussion, though it is of importance for supporting broadband networks. See e.g., Noam (2001).

[68] As of this writing, there was not a viable wireless high speed access technology, though one appeared poised on the horizon, i.e., 802.11g. For the first decade of commercialization all wireless applications, such as wi-fi (802.11b) or the Blackberry (two-way text messaging) used slow speeds.

[69] See, for example, Hurley and Keller (1999), National Research Council (2001, 2002) or Crandall and Alleman (2002) for analysis of last mile issues.

access and "always on," and, at the same time, it is typically priced much more cheaply than a T-1 line. In comparison to dial-up use, broadband access enables better applications and more convenient service, enough benefit for some users to justify paying a price higher than for dial-up.[70] As the volume and complexity of traffic on the Internet increases dramatically each year, the value of these features becomes larger. Furthermore, broadband access has appeal to the vendors. It is a distribution mode that enables providers to offer a wider range of bundled communications services (e.g., telephone, email, Internet video, etc.).

Broadband infrastructure favors urban areas because the technology for broadband is more expensive in less dense locations. DSL technology can only extend 18,000 feet (at most) from the central office switch.[71] The radius is noted to be closer to 12,000 feet for high-quality, low-interruption service. Therefore, one should expect that those living outside this radius will more likely suffer from lack of DSL service. Cable systems are also cheaper in high-density areas for the simple reasons that Cable is expensive to lay in low-density areas (Esbin 1998, Crandall and Alleman 2001).

A provider of broadband experiences high costs when deploying new services. Lack of preexisting infrastructure is to blame. It is expensive to do as a *de novo* investment, because it requires laying direct lines from trunks to a central switch, which involves the expense of dug-up streets, repeaters, and right-of-way permits if the path has no precedent. This type of investment was so expensive in the mid 1990s that only businesses ever incurred the costs—and usually only in a central city or along an existing trunk line, where the costs were lower. Remote rural establishments simply could not afford the fixed private lines that were needed to deliver high-speed data link-ups.

Broadband to the home is also expensive. It is an expensive retrofit on top of either a telephone network (in the case of DSL) or a cable system (in the case of cable modems). For DSL, the switch and lines must be of high quality and outfitted with appropriate software. For cable, the entire cable system needs an upgrade. These expenses arise solely from engineering requirements, so they arise even without the regulatory distortions imposed by state and federal regulators.

To be sure, the costs of deploying DSL are not solely a function of engineering costs. Regulatory rules for DSL can alter the incentives to provide it. Local telephone firms faced an additional "opportunity cost". Demand for additional telephone lines can be quite profitable, but demand for DSL cannibalizes them. In this setting the resale rules, installation expenses, and other regulatory constraints can raise or lower the net returns to providing DSL at incumbent local exchange carriers. It is difficult to make anything other than this general statement, since the regulatory rules for DSL varied considerably across states.

[70] As broadband began to diffuse around 1999/2000, it typically sold for $40 to $60 a month. In comparison, a typical AOL dial up account sold for $23 a month, and plenty of other firms offered cheaper alternatives.

[71] See National Telecommunications and Information Administration (2001) or U.S. Department of Agriculture (2000), for an overview. See Crandall and Alleman (2002), for many studies of broadband deployment and demand.

5.2 The Empirical Evidence

Was there uneven availability across the country? The FCC estimates that high-speed subscribers were present in 97% of the most densely populated zip codes by the end of 2000, whereas they were present in only 45% percent of the zip codes with the lowest population density (NTIA 2002). It also estimates that one-quarter to one-third of U.S. consumers cannot receive this service. Only 90% of the United States has access to cable systems. Most of those who are not reached are located in rural areas. Of those reached, many of those systems require costly upgrades before Internet service is a viable business.[72] This bias can be seen in surveys of use. As of September 2001, 19.1% of Internet users had cable modems and DSL. That percentage could be partitioned into central city, urban, and rural rates of 22.0%, 21.2 % and 12.2% respectively. [73]

Research on the diffusion of broadband, almost by definition, must examine diffusion at a very fine level of granularity, such as at the block or zip code or some other neighborhood-related level of granularity. For example, Lehr and Gillett (2000) look at the very early diffusion of broadband to residential areas (primarily in 1998). They compiled a database consisting of communities in the United States where cable modem service is offered and linked it to county-level demographic data. They find that: broadband access is not universal. Only 43% of the population lives in counties with available cable modem service.[74] This is quite a low number for their methods, which give a county credit if even a small part of the country is being served. Such access is typically available in counties with large population, high per capita income, and high population density; and there is a notable difference in strategy for cable operators with some being more aggressive than others.

In a very data-intensive study, Gabel and Kwan (2001) examine deployment of DSL services at central switches throughout the country, providing a thorough census of upgrade activity at switches. They examine providers' choice to deploy advanced technology—to make broadband services available to different segments of the population. The crucial factors that affect the decision to offer service are listed as (1) cost of supplying the service, (2) potential size of the market, (3) cost of reaching the Internet backbone, and (4) regulations imposed on Regional Bell Operating Companies.[75] They find that advanced telecommunications service is not being deployed in low-income and rural areas.

Prieger (2002), using very comprehensive data, examines the availability of broadband for both cable modems and DSL. He looks for evidence of red-lining, that is, where broadband carriers avoid areas with high concentration of minority or low-income households. He finds little evidence of such red-lining

[72] See, for example, U.S. Department of Agriculture (2000).
[73] One should note that the rate of 22.0% for central cities is likely biased upward due to the presence of universities in the centers of many cities.
[74] They point out this population is actually closer to 27% (as was stated by Kinetic Strategies), but explain that their data is not fine enough to show this measurement.
[75] Data was obtained concerning wire centers; also data on DSL and cable modem service availability was collected via web sites and calling service providers. They supplemented it with Census data.

based on income or against neighborhoods with a concentration of black or Hispanic populations, though the evidence is mixed with respect to Asian or Native American populations. As with others, he finds that broadband is more available in large markets, areas with higher educated populations, and, interestingly, areas with higher Spanish language use. The presence of a Bell company also increases availability, while inner city and rural locations diminish it.

Grubesic and Murray (2004) examine much the same data, but go into deep detail in a few cities, using the closer examination to uncover the drivers of differences from one city to another. They find that density predicts availability if one compares all neighborhoods across the country with each other. But looking inside any given city, the predictors of provision are more varied and not necessarily a function of density. Instead, the wealth and income of the area's residential population is key to understanding demand.[76] Their findings echo Grubesic and Murray's (2002) study of differences in DSL access for different regions in Columbus, Ohio. They observed that DSL access can be quite inhibited because some of the high income areas are also low-density. This supports one of their counterintuitive findings—DSL is less available in some high-income low-density areas, such as Franklin County, Ohio, than in some lower income, high-density areas (See also Grubesic, 2003).

In summary, the spread of broadband service has been much slower and much less evenly distributed than that of dial-up. This is not a surprise once their basic economics is analyzed. The broadband ISPs find highly dense areas more profitable due to economies of scale in distribution and lower expenses in build-out. Moreover, the build-out and retrofit activities for broadband are much more involved and expensive than what was required for the build-out of the dial-up networks. So within urban areas, there is uneven availability. Thus, even before considering the impact of geographic dispersion in demand, the issues over the cost of supply guarantee that the diffusion process will take longer than dial-up ever did.

5.3 The Regulation of Broadband Suppliers

Every city in the United States has at least one incumbent local telephone provider. The deregulation of local telephony was an attempt to increase the number of potential providers of local voice services beyond this monopoly incumbent, and in so doing, increase the competitiveness of markets for a variety of voice and data services. Deregulation became linked to the growth of broadband in two ways. First, these rules shaped the cost and price of providing broadband. Second, deregulation altered the comparative costs of those who supplied local data services, primarily in urban areas.[77] It was easy to anticipate that these regulations would shape who profited from the diffusion of broadband,

[76] See also Grubesic (2004) for a detailed study of the demographic make up of such areas.
[77] See Woroch (2001) for a comprehensive review of the literature. This section is necessarily focused on the issues that shaped the Internet infrastructure market. See, also, Crandall (2001), Crandall and Sidak (2002) or Greenstein and Mazzeo (2003) for exploration of the consequences of deregulation for entrant's behavior.

though, at its outset, it was an open question what those effects would be. An even bigger open question concerned whether these rules shaped the speed with which broadband diffused.

The new competitor for the deregulated network is called a Competitive Local Exchange Company, or CLEC for short. No matter how it is deployed, every CLEC has something in common: each offers phone service and related data carrier services that interconnect with the network resources offered by the incumbent provider (e.g., lines, central switches, local switches). In spite of such commonalities, there are many claims in the contemporary press and in CLEC marketing reports about CLECs that these differences produce value for end users. In particular, CLECs and incumbent phone companies offer competing versions of (sometimes comparable) DSL services and networking services.

Something akin to CLECs existed prior to the 1996 Telecommunications Act. These firms focused on providing high-bandwidth data services to business. After its passage, the firms grew even more. And CLECs quickly became substantial players in local networks, accounting for over twenty billion dollars a year in revenue in 2000.[78] More to the point, CLECs became the center of focus of the deregulatory movement. Many CLECs grew rapidly and often took the lead in providing solutions to issues about providing the last mile of broadband, particularly to businesses and targeted households. In addition, many CLECs already were providing direct line (e.g., T-1) services to businesses (as was the incumbent local phone company).

The incumbent delivered services over the switch and so did CLECs. In recognition of the mixed incentives of incumbents, regulators set rules for governing the conduct of the transactions. As directed by the 1996 Telecommunications Act, this included setting the prices for renting elements of the incumbent's network, such as the loops that carried the DSL line.[79] These rental prices were the subject of considerable controversy, as were the precise rules that governed the obligation of the incumbent to the CLEC.[80]

For this review, the key question is: Did the change in regulations shape the geographic distribution of Internet access across the United States? The answer is almost certainly, Yes, at least in the short run, though the answer is more ambiguous in the long run. By the end of the millennium the largest cities in the United States had dozens of potential and actual competitive suppliers of local telephone service who interconnected with the local incumbent. By the end of 2000, over 500 cities in the United States had experience with at least a few competitive suppliers of local telephony, many of them focused on providing related Internet and networking services to local businesses, in addition to telephone service (New Paradigm Research Group, 2000). This opportunity extended to virtually all cities with a population of more than one-quarter million, as well as to many cities with a population under 100,000. Very few rural cities,

[78] See New Paradigm Resources Group (2000) or Crandall (2001).
[79] See Gregg (2002) for a review of prices across different states. Also, see Rosston and Wimmer,(2001) for review of the determinants of pricing within states.
[80] See, for example, Crandall and Sidak (2002), Sidak and Spulber (1998), and Spulber and Yoo (2003).

however, had this opportunity except in the few states that encouraged it (Malecki, 2002b, Greenstein and Mazzeo, 2003). So at the outset, the entry of CLECs increased broadband supply somewhat only in urban locations, if it had an effect anywhere.

The entry of CLECs as part of the 1996 Telecom Act had a three-fold consequence. First, it encouraged entry by subsidizing it indirectly. For a time, the compensation of CLECs overlapped with the compensation of ISPs. Some ISPs took advantage of rules for compensating telephone companies, a strategy that lost its profitability when the FCC eliminated this distortion in early 1999.[81] Hence, part of the consequence of this entry was temporary, though briefly important.

Second, some CLECs acted as backbone providers for ISPs, thereby spurring the development of local ISP business and allegedly lowering costs for many ISPs. It is unclear how important this was for the geography of the Internet, since many ISPs would have offered service even without the presence of CLECs. The speculation has a basis in fact. While reciprocal compensation may not have influenced the original location of ISPs very much, it had an enormous influence on the profitability, growth and survival of CLECs and that must have had some influence on subsequent ISP development (even after the FCC order that reduced reciprocal compensation).[82]

Third, it is commonly alleged that CLECs spurred the growth of new Internet services in some areas where other potential providers were slow to deploy it. In some cities, at the outset of the Internet, some local telephone companies were reluctant to diffuse DSL services to potential clients, because DSL would cannibalize existing data-traffic business that used T-1 lines, which was quite lucrative. It is unclear how important this factor was for long run deployment, since many telephone companies entered dial-up and DSL businesses as the demand for the Internet grew, and many would have done so eventually without competitive spurs. Hence, this hypothesis requires more careful statistical testing examining the effect of local regulatory stringencies on incumbent provisioning of new services.[83]

It should be added that the 1996 Telecom Act also aided the profitability of cable modem suppliers – in some views this provision of the Act also subsidized the diffusion of broadband to homes. The Act explicitly allowed cable firms to retain their special status outside of the interconnection agreements inherent in

---

[81] See Crandall (2001) for a discussion of the way ISPs took advantage of telephone policies for reciprocal compensation payments. Also see FCC docket No. 99-38, *Implementation of the Local Competition Provisions in the Telecommunications Act of 1996, Inter-Carrier Compensation for ISP-Bound Traffic*, released February, 26, 1999.

[82] If anything, regulatory decisions for reciprocal compensation of CLECs encouraged CLEC entry, which also partly encouraged ISP entry through interconnection with CLECs. Although this is important to incumbent local exchange carriers, one should not exaggerate its effect too much. The scale of this phenomenon grew tremendously in the period between 1996 and 1999, but ISP entry started well before then and continued afterwards until the dot-com bust. Moreover, since CLEC entry was primarily concentrated in dense urban areas, much of this effect was felt in urban areas, which would have had a great deal of ISP entry even without this implicit subsidy to CLECs.

[83] For a beginning on this topic, see Mini (2001) and Koski and Majumdar (2002).

common carrier regulation. Many homes in the United States have a cable line running past it. Internet service to homes through cable lines became a second avenue through which the Internet diffused to homes. Since there is a preexisting urban bias to the deployment of cable systems, this mode of delivery necessarily favored urban areas.[84]

There was considerable variety of experience around these trends. Zolneirek, Eisner and Burton (2001), followed by Malecki (2002b) and Greenstein and Mazzeo (2003), hypothesize that some basic economics lay underneath the manifest patterns: the need to cover these fixed costs limited the number of entrants in specific locales. That is, because of the presence of fixed costs in each location where they provided services, CLECs required a sufficient number of customers to generate revenue in excess of their recurring expenses, given their operating profit margin. As a result, cities varied in their ability to support entrants.[85] Large cities generated sufficient revenue to support more entrants better than small cities did. Large cities receive the greatest number of entrants and small cities the least. These patterns were linked to such things as population size, working population, income, and historical patterns of entry (favoring the largest metropolitan areas).[86] The patterns were also linked to such factors as local regulatory stringency, the identity of the incumbent,[87] the difficulty of managing co-location facilities with a hostile or uncooperative ILEC, and the cost of renting local unbundled network elements from the ILEC (subject to regulatory review).[88]

Unfortunately for suppliers, there were finite limits to how much demand was potentially there for their services. The limits became apparent within a few years of the passage of the Act. The residential market never became particularly large for CLECs. Most CLECs became focused on satisfying business demand, where the margins were higher for basic services and the CLECs could tailor their offerings to complex demands. Even then, some CLECs did not realize revenues sufficient to cover the debts incurred building their facilities and marketing their new services. Though the market revenue of CLECs continued to grow, there was a downturn in the specific prospects of many CLECs. This

[84] This is a long and complex topic. See, for example, Hausman, Sidak and Singer (2001), Crandall and Alleman (2002) or National Research Council (2002) for an overview.

[85] This arises if there are decreasing returns to scale. It can also arise if there are increasing returns to scale, but cournot competition or some sort of differentiation between competitors.

[86] Since the facilities-based CLECs must make capital investments in equipment to link their customers, cities with more geographically concentrated residential neighborhoods and business centers may provide CLECs with customers that are less expensive to serve.

[87] Regulators often have different rules for each incumbent carrier within its state. These rules apply to all the areas within that state where the particular incumbent operates (see Abel and Clements, 2001). Specific provisions in the 1996 act required incumbents to provide interconnection access to CLEC competitors; however, the incentives to comply with this directive differed across firms. Specifically, Regional Bell Operating Companies (RBOCs, for short) that wanted to enter the market for long-distance services were precluded from doing so until regulators were satisfied that the RBOCs had been sufficiently cooperative with CLECs attempting to interconnect and provide service in their local areas (see Shiman and Rosenworcel 2002).

[88] The rental rate of a local loop typically differs across several density zones within each state (Gregg, 2002).

downturn paralleled the downturn among backbone providers and ISPs.

Overall, 2000 was the last year in which there was a consensus of optimistic forecasts about CLEC entry and their expansion. Much of the early enthusiasm was affiliated with anticipated growth in Internet transport or TCP/IP data traffic, development of DSL connections, or affiliated hosting or networking services. It was also affiliated with geographic expansion. By 1999, there was no question that the major cities, such as New York, Chicago or San Francisco, could support some CLECs. But would CLECs spread to medium-sized cities and smaller locales? As it turned out, the answer was yes, but this was harder to do than was initially thought. With low demand in 2000 and 2001 suppliers reduced their growth plans. Many, especially small CLECs or those dependent on growing demand for broadband and a cooperative ILEC to sell DSL, exited altogether.

Any conclusion about this topic comes with caution. The long-term shape of these market actors is currently unclear because the regulatory environment for CLECs is changing. Currently, the FCC is in the midst of rewriting the rules for the rental of unbundled network elements devoted to high-speed data services from incumbent local exchange companies. Courts are in the midst of ruling on the legality of the rewrites being done at the FCC. It is not yet clear how these rule changes will alter the role of CLECs in the geographic development of the broadband markets.

In summary, the era after the passage of the Telecommunications Act of 1996 was a period of uncertainty and flux in the regulatory environment. The rules and regulations certainly shaped the distribution of profits from the supply of broadband. These regulations *might* have speeded up the diffusion of broadband in some urban areas in some states for a short time, especially in the late 1990s, if at all. As yet, there is not yet much evidence these regulations altered the long run geographic properties of broadband – its urban bias. Altering this conclusion requires evidence that one type of firm diffuses broadband faster than another, evidence that no researcher has conclusively presented yet.

5.4 Summary

Economic factors shape the geographic diffusion of broadband. Both DSL and cable Internet is cheaper to deploy in high density areas, so it will continue to be an urban technology. Moreover, the build-out and retrofit activities for broadband are onerous and expensive, so within urban areas there is uneven availability.

The future does not look much different. Broadband will continue to be an urban technology without dramatic changes in regulatory rules or government subsidies. It should also continue until the day that a viable wireless high-bandwidth technology deploys or until the day that someone invents a retrofit to another pre-existing set of lines, such as electrical lines.

Regulation can alter the distribution of rents, by making it easier for one party or another to sell and operate high-speed access. I conclude that there is no conclusive evidence yet that the regulations that affect rent distribution have not had large consequences for the geographic dispersion of broadband.

6. The Location of Business Internet Infrastructure Services

As with many other General Purpose Technologies, advances in frontier technology are only the first step in the creation of economic progress.[89] The next step involves the development of complementary services by economic agents. These developments typically need time, invention and resources before outcomes are realized. This principle applies with particular saliency to the Internet, a malleable technology whose form is not fixed across locations. To create value, the Internet must be embedded in investments at firms and households that employ a suite of communication technologies, TCP/IP protocols and standards for networking between computers. Often organizational processes also must change along various points of a distribution chain delivering services to end users.

As there are several ways to look at the features of that chain, it is no surprise that this topic has invited a variety of perspectives. For the sake of brevity, this review will forego comprehensiveness and focus on understanding the geographic features of the layers next to physical infrastructure. Broadly speaking, many business *users* of Internet infrastructure are also *providers* of Internet infrastructure in a different form, usually related to electronic commerce. Even with this narrowing of the topic, many issues arise about its geographic distribution. In these layers are found a wide variety of businesses and applications.

Research on the relationship between location and private investment in Internet infrastructure divides between two kinds of inquiries that operate on two distinct timescales: First, how did existing business establishments react to the diffusion of the Internet? Second, once the Internet became available, how did firms reorganize or relocate in response? Research has begun to make progress on the first question, because there has been sufficient time to observe the first layers of adoption after commercialization. In contrast, research has not made much progress on the second question, because reorganization and relocation of economic activity tends to occur slowly and only over multiple decades.

Befitting a young research topic, the open questions are still quite basic:

1. What are the differences in the predominant patterns of business infrastructure investment between urban and rural locations?
2. Are there any measurable consequences for location of economic activity resulting from differences in investment in Internet infrastructure?
3. What is the evidence, if any, that the diffusion of Internet infrastructure is reshaping the factors supporting urban agglomeration?

6. 1. The Geography of Private Investment in IT—Overview

The Internet was a malleable technology when it first commercialized in 1992. It needed to be adapted for commercial use. Adaptation was necessarily a

[89] For more on the theory of general purpose technologies, see Bresnahan and Trajtenberg (1995). For discussion about the comparison of the economic valuation of new Internet technology to users in different regions, see Bresnahan and Greenstein (2001).

local economic activity, resulting from the combination of the local demands of business establishments and the supply constraints of markets for Internet technology infrastructure and services. More to the point, the same technological opportunity – that is, the commercialization of the Internet across the United States – did not result in the same commercial experience for all establishments in all locations.

As the Internet commercialized, the vast majority of business establishments were faced with a decision about how to react to the availability of new capabilities. As it turned out, American businesses reacted with the largest growth rates in investment in IT in the history of the United States. Stocks of information technology capital grew at a 20% annual growth rate from the end of 1995 to the end of 2000.[90] By 2000, computer hardware and software stocks had reached $622.2 billion.[91] The majority of this investment was affiliated with enabling business applications. In 2000, for example, total business investment in IT goods and services was almost triple the level for personal consumption of similar goods.[92]

The level of these investment flows is immense and so is the variance across locations.[93] In some locations, the Internet has been adopted across all facets of economic activity, while in other locations adoption is not widespread. What explains this variance? Two distinct views have become prominent.

One view argues that Internet technology requires infrastructure and support services, which are more readily available in urban settings. It forecasts that businesses in urban settings use and deploy Internet applications more frequently than do similar firms in rural settings. It also argues there was little exceptional about the economic geography of the Internet. As with previous frontier IT, most of the productivity benefits from these investments accrued to urban businesses. A contrasting view argues that Internet technology was exceptional, different from all the IT that came before it. According to this view, the Internet decreases coordination costs within firms and between firms, which reduces the importance of distance. Internet technology dramatically reduces the costs of performing isolated economic activity, particularly in rural settings, even when deployment costs are high.

A related debate analyzes how the diffusion of the Internet alters the roles of cities and urban agglomeration in economic life.[94] Again, there are two related viewpoints. One view sees a diffusion process that enhances already existing advantages to urban areas or central cities. The contrasting view observes a diffusion process that diminishes the advantages of urban areas in favor of locations that had previously been considered isolated. The open question

[90] This includes computer hardware, computer software, and communications hardware and instruments. See Price and McKittrick (2002) or Henry and Dalton (2002). The growth rates are even higher if communications hardware and instruments are excluded.
[91] These are constant (1996) dollars. See Henry and Dalton (2002).
[92] For 2000, estimated personal consumption of IT goods and services was $165 billion. For business it was $466 billion. See Henry and Dalton (2002).
[93] For an extensive description of this variance, see Forman, Goldfarb and Greenstein (2002).
[94] For a general entry into these issues, see, e.g., Cairncross (1997), Castells (2002), Kotkin (2000) or Dodge and Kitchin (2001).

concerns the complementarity or substitutability between the Internet and urban agglomeration?[95]

These two views have deep roots in long-standing debates about the location of economic activity. These led to multiple potential explanations for why advanced IT located in urban areas. These emphasize the costs of adoption, such as (1) availability of complementary information technology infrastructure, such as broadband services, (2) labor market thickness for complementary services or specialized skills, and (3) knowledge spillovers.[96] One other explanation emphasizes that the types of firms found in urban areas are not random. That is, historically IT-friendly establishments may have sorted into areas where costs have previously been low for precursors to Internet technology.

The opposing view emphasizes that establishments in rural or small urban areas derive the most benefit from overcoming diseconomies of small local size. That is, Internet technology substitutes for the disadvantages associated with a remote location. There are several reasons why this may be true. For one, use of Internet technology may act as a substitute for face-to-face communications.[97] Common examples are email or instant messaging. Second, establishments in a rural area lack substitute data communication technologies for lowering communication costs. Third, advanced tools such as groupware, knowledge management, web meetings, and others also may effectively facilitate collaboration over distances. Some tools enable simultaneous changes to electronic documents by users in multiple locations. Moreover, supply chain management software enables electronic communication of data that would be costly to transmit via phone or through the mail.

Properly answering these questions requires extensive documentation of business use of the Internet. While there has been considerable empirical research on the use of the Internet at households, there has been much less systematic empirical research on the geographic features of Internet technology use by business. Research addressed one of two related goals. First, what is the variance between locations in the average use of Internet technology for some purpose? Second, what is variance in the marginal contribution of the location – instead of some other factor – to the observed outcome?

6.2 Empirical Evidence on Domain Names

As the first probe into these issues, several researchers measured the Internet content produced across the United States. For these studies, researchers find the location of each firm with a dot.com Internet address and plot it. Domain names are used to help map intuitive names (such as www.northwestern.edu) to the numeric addresses computers use to find each

[95] See, for example, Glaeser and Gasper (1997), Kolko (2002), Sinai and Waldfogel (2000).
[96] These are closely related to the three major reasons given for industrial agglomeration (e.g., Marshall 1920; Krugman 1991).
[97] Other authors (e.g., Gaspar and Glaeser, 1998) have argued that improvements in information technology may increase the demand for face-to-face communication. In other words, they argue that IT and face-to-face communication may be complements. The implication of this hypothesis is that commercial establishments relocate to urban areas in reaction to technical change in IT.

other on the network.

Host site counting presents challenges as a measurement of economic activity. It cannot account for the common practice of not physically housing Internet-accessible information at a firm's physical location.[98] So, while it is informative, such evidence is an imperfect way to measure the geographic dispersion of business activity that acts as Internet infrastructure.

If anything, this research has not yet settled on a firm set of hypotheses to test. Authors have tried to use this data to examine classic issues in geography and communications. As Townsend (2001a) argues, for example, with the rise of the Internet, the notion of a global city was questioned, and so too were the two dominant theories on the relationship between telecommunications and urban growth. One theory forecast the centralization of decision-making in "global" cities, the other forecast wholesale urban dissolution.[99] As yet, there is little evidence for either theory; he argues that we neither see evidence of dominant global cities nor do we see total dissolution. As noted early, it is unsurprising that this debate remains open. It requires evidence that can only come about when establishments relocate, a process that proceeds slowly.

In a pioneering set of studies, Moss and Townsend (2000a, 2000b) analyze the growth rate for domain name registrations between 1994 and 1997, the early years of Internet development. They find that "global information centers" such as Manhattan and San Francisco grew at a pace six times the national average, while global cities such as New York, Los Angeles, and Chicago grew only at approximately one to two times the national average.

Examining a mildly later set of registrations, Zook (2000) also finds that San Francisco, New York and Los Angeles are the leading centers for Internet content with regard to absolute size and degree of specialization. Degree of specialization is measured by relating the number of dot.com domains in a region relative to the total number of firms in a region to the number of dot.com domains in the United States relative to the total number of firms in the United States.[100]

Kolko (2000) also analyzes such data, but is the first to provide an economic framework in which to understand the outcomes in central and peripheral locations. He considers whether the Internet enhances the economic centrality of major cities in comparison to geographically isolated cities. He argues, provocatively, that reducing the "tyranny of distance" between cities does not necessarily lead to proportional economic activity between them. That is, a reduction of communications costs between locations has ambiguous predictions about the location of economic activity in the periphery or the center. Lower costs can reduce the costs of economic activity in isolated locations, but it can also enhance the benefits of locating coordinative activity in the central location. As

[98] Also, it may be unable to differentiate between various types of equipment. See also Warf (2001) for similar results to those discussed herein, which suggests that there is robustness in some of the observations.
[99] See also Gorman (2002).
[100] Townsend (2001a) argues that the leading cities in total domain names identify the presence of many content providers. He also finds that New York and Los Angeles top the list. He makes much of his findings about Chicago—normally considered along with New York and L.A. as a global city— which only ranks a distant fifth, far behind the two leaders.

with other writers, Kolko presumes that coordinative activity is easier in a central city where face-to-face communications take place (also see e.g., Glaeser and Gaspar, 1997).

Kolko's findings do not settle the question, but do raise intriguing issues. As with other researchers, he documents a heavy concentration of domain name registrations in a few major cities. But his framing motivates him to examine the role of proximity versus remoteness. In particular, he also documents extraordinary per capita registrations in isolated medium-sized cities. He argues that the evidence supports the hypothesis that the Internet is a complement, not a substitute for face-to-face communications in central cities. He also argues that the evidence supports the hypothesis that lowering communication costs helps business in remote cities of sufficient size (i.e., medium-sized, but not too small).

Zook (2001) also examines the locations of the dominant firms in e-commerce. He finds the location of the top Internet companies on the basis of electronically generated sales and other means. His analysis shows San Francisco, New York, and Los Angeles as dominant in e-commerce, with Boston and Seattle not far behind. Many Midwestern cities such as Detroit, Omaha, Cincinnati, and Pittsburgh, as well as many cities in the South are lagging behind those in other parts of the country. Zook (2000) measures the overall size of e-commerce in a region and the E-commerce specialization of a region relative to the number of Fortune 1000 firms headquartered there. He points out that the cities that are major business centers are not necessarily the ones leading in e-commerce adoption. Despite this new approach, once again, he finds a similar ranking of cities and regions.

Another approach followed by Townsend (2001) analyzes domain name density. This is a ranking of metropolitan areas according to domain names per 1000 persons. Among Los Angeles, New York, and Chicago, only Los Angeles ranks among the top twenty (17^{th}). The full ranking over domain name density indicates that medium-sized metropolitan areas dominate. Many global cities rank highly, as do many medium sized areas, whereas many small metropolitan areas show very low levels of Internet activity. Though statistically simpler than Kolko's analysis, this evidence is consistent with Kolko's view that sufficiently large medium cities benefited from the Internet's diffusion.

A related set of studies analyzes registrations at the level of the neighborhood. This type of study indicates that business needs determine the location of registration activity. A detailed analysis of registrations from the early commercial boom time found that adoption in Manhattan (Moss and Townsend, 1997) and San Francisco (Zook, 1998) was centered on central business districts in those cities. Grubesic's (2002) detailed examination of Ohio – at the zip code level in 1999 – highlights the considerable differences across Ohio's diverse areas. Once again, urban centers of the state have higher activity than the suburbs and the rural counties. At this level of detail it is also apparent that universities can spur further registrations above and beyond that predicted merely by economic factors.

There is a sense in which these findings about domain name registration are not a huge surprise, but it does suggest open questions to explore. Further work needs to be done to understand the links between the location of activity within cities and the concentration of new venture formation, the commuting patterns of highly skilled and educated labor, and the location of other facets of business in dense and non-dense locations. Related, now that the dot-com crash has led to many bankruptcies and exit, it is interesting to know whether the geographic distribution of domain names was informative about new ventures only or also about durable business location practices.

6.3 Evidence on Urban and Rural Business Use of Internet Technology

Another challenge for research arises from confusion about the uses of Internet infrastructure in business. Forman, Goldfarb and Greenstein (2003a, 2003b, 2003c) observe there were many purposes for the Internet in business, and show how these different purposes inform studies of the geographic diffusion of infrastructure.

They contrast two purposes, one simple and the other complex. The first purpose, labeled *participation*, relates to activities such as email and web browsing. This represents minimal use of the Internet for basic communications. The second purpose, labeled *enhancement*, relates to investment in frontier Internet technologies linked to computing facilities. These latter applications are often known as *e-commerce,* and involve complementary changes to internal business computing processes.

The economic costs and benefits of these activities are quite distinct, but not highlighted by previous research. Adaptation costs are relevant to the adoption decision for enhancement and largely negligible for participation. The costs of installing enhancement will be more sensitive to increases in density than will participation, while the (gross) benefits from participation will be higher in rural areas than in dense areas. The open question concerns the sensitivity of the "net benefits" of use to different geographic locations.

Forman (2002) was the first paper to put such distinctions to use. Forman examines the early adoption of Internet technologies at 20,000 commercial establishments from a few select industries. He concentrates on a few industries with a history of adoption of frontier Internet technology and studies the microeconomic processes shaping adoption. He finds that firm with dispersed establishments were more likely to adopt Internet technology for purposes of participation. Rural establishments were as likely as their urban counterparts to participate in the Internet and to employ advanced Internet technologies in their computing facilities for purposes of enhancement of computing facilities. He attributes this to the higher benefits received by remote establishments, which otherwise had no access to private fixed lines for transferring data.[101]

Seeking to generalize this type of study to other industries, Forman, Goldfarb and Greenstein (2003a) examine all private non-farm business

[101] See, also, Premkumar and Roberts (1999), who identify the state of use of various communications technologies and the factors that influence the adoption of these technologies in rural small businesses in the United States. The authors collected data on seventy-eight organizations in a structured interview process, and their results indicate the importance of many factors. They also express concerns that rural businesses may lack access since Internet infrastructure tends to follow a similar pattern as the interstate road system.

establishments with 100 or more employees at the end of 2000. This includes a wide array of establishments from manufacturing, service, finance, and for-profit activities. They show that adoption of the Internet for purposes of participation is near saturation in most industries, with some marginal differences for locations. Their findings for enhancement contrast sharply. Establishments in MSAs with over one million people are one and a half times as likely to use the Internet for enhancement than are establishments in MSAs with less than 250,000 people.

They conclude that rapid diffusion in participation did not necessarily imply rapid diffusion in enhancement, speculating that diffusion of enhancement will follow a more traditional path than participation. That is, it will take time, innovation, and resources before economic welfare gains are realized. They also concluded that that Internet use in business is widely dispersed across geographic regions. They conclude that research focused on concentration or digital divides – heretofore a central concern of the literature on Internet geography—is a misleading basis for formulating regional economic policy about Internet use in business.

Why do some regions lead and others lag? Forman, Goldfarb and Greenstein (2003c) focus on two factors, the marginal contribution of location to the propensity to use the Internet and the marginal contribution of an establishment's industry. They find that the variance of experience is quite narrow within large MSA – there is little difference between the best area, San Jose, and the worse, Las Vegas. Moreover, the difference within region widens in comparison of large, medium, small and rural MSAs. Hence, the competitive differences between establishments of the same type are greater for smaller metropolitan areas. The difference between the average large MSA area and the worse small MSA area is also great. In extreme situations, location can make a large difference by itself. For example, establishments in a small and medium MSAs, who already are suffering from slow growth, will typically have low adoption rates of Internet technologies.

Overall, however, an establishment's use of advanced Internet technology is mostly explained by the industry for which it produces. They conclude that the (pre-existing) distribution of industries across geographic locations explains much of the differences in average rates in enhancement between locations. Large urban areas lead in the use of advanced Internet applications because the industries that "lead" in advanced use of the Internet tend to disproportionately locate in urban areas.

This conclusion highlights that some industries are more information intensive than others, and, accordingly, make more intensive use of new developments in information technologies, such as the Internet, in the production of final goods and services. Hence, the geographic dispersion of modern infrastructure will partly depend on whether the information-intensive industries tend to be more geographically concentrated than the less information-intensive industries. Heavy Internet technology users have historically been banking and finance, utilities, electronic equipment, insurance, motor vehicles, petroleum refining, petroleum pipeline transport, printing and publishing, pulp and paper, railroads, steel, telephone communications and tires (Cortada, 1996).

Forman, Goldfarb and Greenstein (2003b) show that establishments from industries with a high propensity to use advanced technology tend to locate in urban areas. They argue that several factors contribute to this persistence of demand for advanced IT at the same establishments in the same industries one decade after another. First, firms are incremental in their approach to investment, compromising between the benefits of frontier and the costs of keeping an existing process, picking and choosing among those new possibilities that make the most sense for their business. This is consistent with Forman's (2002) finding that installed base of hardware and software applications played a major role in shaping organizations' early decisions to invest in the Internet.

Second, consistent with Cortada (2004), they argue that investment in innovative IT is directed toward automating functional activity or business processes within an organization, such as accounts receivable, inventory replenishment or point of sale tracking. If there is stability of the types of economic activity going on within organizations, then this stability shapes the demand for innovative IT, enhancing the same activities decade after decade. Related, most organizations examine other firms with functions similar to their own and benchmark their own processes against them. Such activity increases the incentives of lead firms to emulate other organizations in the same industry, either close competitors or those with similar supply chains.[102]

A related facet of this topic concerns the geographic concentration of labor markets. Because Internet use in business employs highly skilled and educated workers, there are open issues about how the geographic distribution of skilled labor shapes the development of the next layer of Internet infrastructure.[103] Related studies of the labor markets for Internet services are still in their infancy. This type of research involves analyzing technically adept labor markets, which are characterized by high mobility between firms.

In a particularly striking paper in this vein, Jed Kolko (2002) examines the Internet in light of previous evidence that technology-intensive industries exhibit slower employment convergence than other industries. In this context, *convergence* refers to the tendency of an industry to become more uniformly distributed geographically. [104]

Kolko argues that it is not necessarily information technology usage that slows convergence but that the industry hires more educated workers, which causes clusters to persist. Although convergence appears to be slower for Internet technology industries, this conclusion is deceptive. When compared to industries that hire educated workers, Internet technology actually speeds up convergence. Kolko makes a theoretical case as to why certain aspects of the

[102] For a novel attempt to statistically model these choices, see Fitoussi (2004), who examines the choice over the location of Customer Research Management functions for Fortune 1000 firms.

[103] For a discussion of the role of skilled labor within organizations using advanced IT, see Bresnahan, Brynjolfsson and Hitt (2002), Chapple and Zook (2002), or Feldman (2002).

[104] For earlier related related works about the geography of the computer industry, see Beardsell and Henderson (1999) or Dumais, Ellison and Glaeser (2002). In both of these works, because of longer time periods, researchers have begun to identify the sensitivity of location decisions to labor market features, and visa-versa.

Internet technology industry might converge quickly, namely, knowledge spillovers are less dependent on location and transportation costs are small. Nevertheless, he finds slow convergence.[105]

Gorman (forthcoming) takes another approach to this topic by focusing on the top forty web integration firms in the US, such as KPMG or Accenture (Formerly, Andersen Consulting). These firms expanded during the dot-com boom, and Gorman proposes to examine their hiring and employment practices at their home and major branch offices. He argues that this is informative about the distribution of skilled IT labor. He finds a strong bias toward major urban centers, such as New York, San Francisco, Boston, Los Angeles, and Chicago. On one level this is not at all a surprise: It suggests that the skilled labor in these large firms concentrate their home and major branch offices in major cities, where they have access to better IT infrastructure as well as other services, such as large airports. On the other hand, that seemingly obvious pattern is Gorman's main point – that, contrary to utopian visions about the irrelevance of location, the skilled labor for IT did not pick up and largely move to a small or medium sized MSA that many people consider desirable. That place could be Madison, Wisconsin or Durham, North Carolina, or Santa Barbara, California – for Gorman's purpose it does not matter. For a variety of reasons the largest urban centers continue to be the central locations for skilled IT labor markets, even during a boom time.

More research of this type is needed to confirm these observations for more recent experience. For example, as the market for skilled IT labor contracted, how did different areas of the country adjust? As the web enables communications over longer distances, which types of skilled labor becomes more mobile, if any? Which functions move most easily between locations, activities that directly build web infrastructure, such as highly skilled programming, or activities that make use of communications, such as call-centers?

6.4 Local Internet and Information Services

Local information services, such as newspapers, are another important economic activity to examine. These activities use Internet infrastructure heavily, but also extend it in new ways for local and national readers.

Chyi and Sylvie (1998, 2001) examine the local and long-distance on-line readership for on-line newspapers. Data was gathered via a survey of sixty-four newspapers with on-line versions. They find that the long-distance market is a substantial sub-market constituting about one-third of the on-line readership, but the local market still outweighs the long-distance market in terms of usage and the on-line newspapers' targeting intention. They stress the importance of recognizing the difference between their local-market operation and long-distance market. This further argues for the view that local newspapers use the on-line versions as supplements for enhancing revenue, rather than as substitutes that cannibalize existing revenue streams.

[105] His data come from the Longitudinal Enterprise and Establishment Microdata file (1989-96) and the Current Population Survey.

Boczkowski (2002, 2003) provides a close analysis of many of the experiments undertaken by on-line newspapers during the late 1990s. He documents many of the motives behind local newspapers developing their on-line divisions, and the consequences that resulted. He shows that most newspapers did not emphasize developing new markets in the sense of finding new customers. Rather, many of these newspapers focused on developing new capabilities that provided new services to their existing customers. This could involve using the interactive capabilities for developing on-line communities or using new media to enhance coverage in ways that were not viable in a print form. The focus on local customers also followed from the commercial motivation inherent in trying to develop new media tied to ad revenue, much of which came from local sources. The general lesson is that the costs and motives of a local information provider shaped the exploration, extension and direction of development of local Internet infrastructure into information services. The commercial motives of the actors directed this technology towards geographically local services.

In a novel approach to the question, Sinai and Waldfogel (2000) examine what users do on the Internet. They inquire whether the Internet may serve as a substitute for urban agglomeration by leveling the consumption of Internet content between large, variety-laden markets and small variety-starved markets. The key issues revolve on whether content on the Internet is similarly attractive to persons in large and small markets. They recognize that it can go either way. If the Internet equally aggregates the interests of many users, then those in isolated locations gain. If the Internet offers local, as well as general, information, then it may not necessarily have a role as a substitute for agglomeration. If local on-line content is sufficiently attractive and more prevalent in large markets, then, in some states of the world, it is possible for the Internet to act as a complement to urban agglomeration. Indeed, they find mixed evidence that the Internet acts as both a substitute and complement for urban agglomeration, depending on the type of individual.

Another line of inquiry examines how the Internet alters old definitions for geographic space and redefines urban living spaces. This line of inquiry arises out of an older set of studies examining the impact of the telephone on aspects of urban economic life, such as commuting.[106] There are related questions to ask about the impact of Internet-enabled mobile applications, such as Wi-Fi (i.e., 802.11b or its descendents) or two-way text messaging (i.e., Blackberry or Palm E-mail) or video-conferencing. Similarly, how do these applications affect the use of public spaces, such as airports and café's, or the travel patterns of particular occupations, such as sales representatives or consulting? This line of inquiry can go in many directions. For a variety of reasons, wireless telephone infrastructure is comparatively ubiquitous across all regions of the US (Gorman and McIntee, forthcoming). An effective mobile application that builds off this infrastructure could quickly become available everywhere.[107]

[106] See, in particular, Kitchin (1998), who discusses the growth of "back-office" operations, teleworking, and overall employment restructuring.
[107] This was one of the facets that helped the diffusion of the Blackberry, for example. It employs

In this vein, Light (1999, 2001) raises questions about how the Internet will alter public life, especially the part of public life taking place in commercial settings. Previously, changes in architecture in cities have been cited as explanations for decline in public life, and scholars are predicting further changes due to the Internet. While people might think the Internet could replace small-town street-corner societies with its virtual communities, Light believes such forecasting is in haste, similar to the way the mall was originally thought to be a replacement for small-town street corners. She points out that the areas of the Internet that act as a location for civic life are generally under strict controls and quite commercialized (on-line communities, specifically). Light does not dismiss concerns that the Internet may erode public life or alter daily patterns of use of public areas, but she does argue that many of these concerns are a bit overblown—as she believes that perspectives on cyberspace are likely to evolve.

6.5 Summary

Has the Internet realized its promise of reducing the importance of location to economic activity? By 2000, participation activities such as email and web browsing had diffused almost everywhere. For complex technologies such as enhancement, in contrast, IT-intensive firms found greater benefits than other firms from pooled resources in large cities.

Does the flow of investment dollars correlate positively with the rankings of location and industries? Were the investment dollars affiliated with the commercialization of the Internet widely dispersed throughout locations and industries in the United States? These questions are important for understanding the impact of the Internet's diffusion of regional growth. They are subject to speculation, and await confirmation with data about investment behavior beyond adoption of infrastructure to support the first generation of e-commerce applications.

Business use of information infrastructure should continue to be a robust research topic. Investment done by private firms rivals or exceeds investments at traditional carriers. Privately owned equipment lives side-by-side w/ the public switch network & quasi-private Internet carrier network in arrangements without precedent. Moreover, these business networks potentially form the basis of regional comparative advantage. This network arises from the actions of multiple investors, potentially responsive to new concerns. As the commercial Internet moves into its second decade, the geographic properties could be quite different from those more familiar in telephony.

AT&T's wireless infrastructure for mobile text-messaging.

7. Summaries of Answers to Motivating Questions

The NSF began to commercialize the Internet in 1992. Within a few years there was an explosion of commercial investment in information infrastructure in the United States. Research on the geography of the Internet seeks to understand the durable economic processes underneath the Internet's growth and seemingly unique commercial experience.

At the outset of commercialization the economic determinants of the commercial Internet were unknown. Upon examination one can see that a lot of the Internet infrastructure was caused by the same familiar economics that have been seen with other communication networks. This review has touched on familiar economic concepts – i.e., the economies of density and scale in operation, the economies of entry into provisions of services with high sunk costs, the economics of retrofitting technical upgrades on existing infrastructure, and the economics of competitive behavior for growing markets. These have shown up in new places with new sets of economic actors, to be sure, but the analysis required borrowing familiar frameworks from existing communications markets, not a fundamental new set of economic precepts. In that vein, below I summarize answers to the questions that motivated the review.

7.1 Why Did Near-Geographic Ubiquity Arise after Commercialization?

The Internet access business was commercially feasible at a small scale and, thus, at low levels of demand. This meant that the technology was commercially viable at low densities of population, whether or not it was part of a national branded service or a local geographically concentrated service. This partly mimicked the academic experience, where the operations were also feasible on a small scale. That statement alone does not capture all the factors at work, however.

Two important factors prevalent throughout the United States made Internet access feasible in a wide variety of organizational forms, large and small. First, entrepreneurial initiative helped small-scale business opportunities thrive. These businesses included those located in many low-density cities, isolated cities, or rural areas that were not being served by the national firms.

Second, entry was inexpensive. Small-scale implementation depended on the presence of high-quality complementary local infrastructure, such as digital telephony, and interconnection to existing communications infrastructure, which was available due to national and local initiatives to keep the communications infrastructure modern. It also depended on the widespread use of flat rate pricing for telephone service in most locales. Thus, the economies of retrofits took over entry decisions, not the economies of building networks *de novo.*

TCP/IP based technologies can alter the trade-off between frontier applications and costs. Going forward, one should expect similar advantages for new applications that build on pre-existing infrastructure. For example, applications that operate within cell-phones (e.g., Real Networks) or PDAs (e.g., Palm-based or Windows CE based) or laptops (e.g., Wi-Fi) appear so promising because these build on top of pre-existing infrastructure.

7.2 Why Did Market Forces Encourage Extensive Growth?

Market forces can customize new technologies to users and implement new ways of delivering technologies. These activities have immense social value when there is uncertainty about technical opportunities and complex issues associated with implementation. As such, this industry offers many examples for illustrating lessons from the economics of market-based experiments.

The ISPs knew about the unique features of the user, the location or the application. This was precisely what was needed to customize Internet access technology to a wide variety of locations, circumstances and users. Removing the Internet from the exclusive domain of NSF administrators and employees at research computing centers brought in a large number of potential users and suppliers, all pursuing their own vision and applying it to unique circumstances. In addition, it allowed private firms to try new business models by employing primitive web technologies in ways that nobody at the NSF could have imagined.

Yet, the location of experiments was necessarily temporary, an artifact of the lack of maturity of the service. As this service matured—as it became more reliable and declined in price so that wider distribution became economically feasible—the geographic areas that were early leaders in technology lost their comparative lead or ceased to be leaders. As such, basic ISP technology diffused widely, to places other than universities, and comparatively rapidly after commercialization.

In contrast, experimentation in the commercialization of the backbone resulted in the build-out ahead of demand. Not only was this a fast build-out, but it also resulted in an oversupply in some locations. As of this writing, it is generally expected that the oversupply of capacity will likely persist for some time, even though demand continues to grow.

The middle of the decade after commercialization also illustrates behavior during times of uncertainty, as well as growth in demand. Firms thought that they were in a race to (1) meet anticipated demand and (2) build capacity to sell to others. These incentives appeared to be good before the backbone was ever built, since it sped up development. After the network was built, these incentives appear to have been too strong, leaving many firms holding assets with no customers. In retrospect, these were imprudent investments. Such regret is standard for markets after a period of exploratory behavior aimed at resolving commercial uncertainty, so the welfare conclusions are necessarily ambiguous.

7.3 Has the Internet Diffused Disproportionately to Urban Areas?

Researchers continue to find two persistent patterns that stand in marked contrast: Some see urban areas being disproportionately favored, others find that the technology has spread ubiquitously.

Three factors help reconcile these contrasting assessments. First, transmission technology displays economies of scale. For the foreseeable future, there will be “big pipes” traveling between “big switches.” Central locations like

Chicago in the middle of the country, New York and Washington, DC in the east, Dallas and Atlanta in the South, and San Francisco and Los Angeles in the West, will continue to have access to big pipes going to it from almost everywhere in the country.

Second, as with many high tech services, areas with complementary technical and knowledge resources are favored during the early use of technology. This process favored growth in a few locations, such as Silicon Valley, the Boston area or Manhattan – for a time at least, particularly when technologies was young. But did it persist? For a wide variety of uses of the Internet, the answer is simply no. The common perception that the use of the Internet concentrated in a few locations is false, so too is the perception that the revolution dramatically altered the use of IT in only a few locations or a small number of industries.

This technology matured into something standardized that could be operated at low cost in areas with thinner supply of technical talent. In the first generation of Internet applications, standardization occurred quickly. Just as the Internet moved quickly to locations distance from Universities, so too the first generation of commercial infrastructure did not concentrate in a few large cities. It diffused to virtually every large city and most medium and small cities. In other words, such local factors did not shape the difference between major metropolitan areas as much as it shaped the difference in experience between large MSAs and a few small MSAs or rural areas – not the majority of such small areas, who did just fine. Only a very qualified statement is consistent with the experience of the 1990s, not the alarmist tone found so commonly in discussions about the digital divide.

Third, the next generation of Internet infrastructure, broadband, is a geographically local technology in most of its commercial forms. Unlike the deployment of dial-up, the economics of this deployment strongly favor areas of urban density. Even within that landscape, the build-out has many facets of uncertainty. In particular, regulatory decision-makers at the national and local level continue to hold key levers in determining whether local telephone firms or differentiated competitors (interconnecting with the local network) are the commercially more profitable organizational form for diffusing and commercializing this delivery mode. Accordingly, forecasting the future in this environment is virtually impossible.

7.4 Is the Internet a Substitute or a Complement for Urban Agglomeration?

In many respects this often-posed question is not posed very well. Urban agglomeration occurs for many reasons, and only a few of these have direct relationship with the use of information technology. Cities serve multiple purposes to their denizens and the diffusion of a new communications technology will alter some of those reasons, while not changing others at all. The diffusion of the Internet alters both the costs of doing some activities in urban centers and the benefits of doing them elsewhere.

The deeper question is really one about comparative regional advantage, and the competition takes place on multiple margins. Relocation can occur in many different ways. Firms may choose between central cities and suburbs, between urban areas and less dense settings, or, for that matter, between a location within the US and an area outside the boundaries of the US.

After a decade of the diffusion of the commercial Internet, it is not obvious that cities have become comparatively less favorable as a location for economic activity. First, cities still retain many of their economic advantages as a location for coordinative activity in comparison to rural areas. Some of this has to do with access to information infrastructure, but most of it arises from access to other city services, such as major airports or transportation hubs, not to mention other shared public goods, such as security services, urban amenities, cultural activities, and so on. Second, cities also retain advantages over rural area due to the thickness of local labor markets for technical talent. For a variety of reasons, cities will continue to be attractive locations for high-technology talent.

I conclude that, for the foreseeable future, urban areas will serve as the home to the establishments from the industries with long-standing information-intensive demands on the frontier. Perhaps a few key industries will now be able to move some activities outside major urban areas, but it is difficult to see much evidence that the movement on this margin will be dramatic. For example, some activities can move abroad, such as telemarketing, while others can be based out of any US location, such as auditing or aspects of tele-medicine, such as reading MRI results. Other financial transactions will also be able to move more easily, conditional on regulatory constraints. That is, the analysis of mobility will take place at this functional level, if at all, and appears not to have broad or sweeping determinants.

The comparison between cities is less obvious. In the comparison of one major city to another (e.g., Los Angeles compared to New York), the diffusion of the Internet probably did not alter the comparative rankings for particular places. Fast communications generates multiple responses. Some activities, such as the financial transactions typical of a corporate headquarters, no longer need to be located at the point of production and can also move to take advantage of complementary factors inside or outside a city, such as land costs, congestion externalities, or the location of complementary services. But where will such headquarters go, to another central city, to a nearby suburb, or abroad? The evidence is sparse, so the answers are still quite speculative.

Further research needs to refine the question. The first effect of a new technology emerges through adoption behavior. As users become accustomed to ubiquitous Internet technology, a second effect emerges. This effect is associated with endogenous location decisions with such infrastructure in place. Whereas adoption behavior can take place on a short timescale – i.e., less than a decade – relocation takes place on a much longer timescale. The full effects on different industries will not be understood until after decades of change.

7.5 Which Policies Mattered? Were These Effects the Intended or Unintended Consequences?

The US federal structure is quite unique by International comparison, and gives rise to regulatory tensions that will not arise elsewhere. More broadly, United States regulatory policy imposed a number of constraints on economic actors that shaped outcomes – in the sense that behavior would have been different had policies been different. With that in mind, it is important to catalogue a few policies that contributed to the Internet's comparative ubiquity across the United States.

Some policies had an unambiguous salutary effect, but were unintended consequences of previous decisions. For example, free local phone calls over a short distance had unintended positive effects on the Internet's adoption and ubiquity, particularly when the commercial Internet was quite young. A similar remark could be made about previous national programs to upgrade all switches to digital technology, even those in rural locations, which was done for reasons more related to universal service and because it enhanced the diffusion of complementary digital technologies, such as the Fax machine. The Internet came along much later and piggy-backed on these upgrades, but would have been difficult without them.

Perhaps the most significant policy – as far as unintended consequences – came from the specter of anti-trust law hanging over communications market structure for equipment and carrier services. Prior antitrust decisions – everything from Judge Green's constraining implementation of the divestiture decree to the FCC orders affiliated with deregulating long distance services and customer premise equipment markets – gave the US its eccentric assembly of incumbent suppliers. There were carriers, such as AT&T, WorldCom, Sprint, and the Regional Bell Operating Companies. There were service providers and equipment firms, such as IBM, Cisco, or Lucent, or, for that matter, AOL and other firms with experience in the bulletin board industry. Said succinctly, this was not the IT market of the 1960s or 1970s, when IBM and AT&T dominated so many design decisions and controlled a high fraction of the distribution of equipment, determining how most business users took advantage of the IT frontier.

The 1990s were an era of widely dispersed technical knowledge. The economic actors of the 1990s approached the commercial opportunities afforded by the Internet with quite different experience and capabilities. In the face of market uncertainty over the basic foundations of costs and demand, this variance produced a variety of market based actions. The basic economics of experiments suggests that a variety of views is healthy in an environment where the experts are uncertain because it leads to rapid exploration of a variety of combinations of otherwise unknown demand and cost conditions. US policy fostered this variety by discouraging concentration of decision making in a small number of firms.

At a broader level, it is clear in retrospect that other regulatory decisions also permitted firms and users to experiment. For example, one set of important rules – dating to the FCC's old decisions for Computer I, II and II – delimited local telephone firm discretion over interconnection between the network and customer premise equipment. These rules eventually resulted in the FCC's mandate for standard physical interconnection between phones and wires, a physical connection that made a PC as easy to hook up to a telephone wire as it is to hook up a facsimile machine to a wire.

A related set of rules delimited incumbent behavior in their relationship with other carriers. These interconnection rules allowed bulletin board providers to interconnect without regulatory oversight or hassle, the type of precedent that allowed the ISP industry to grow rapidly. Related rules provided entrants a measure of protection against exclusionary behavior by incumbents, which permitted backbone providers to deploy sunk assets on a large scale and with confidence that they would go to use. Such rules enabled behavior that developed multiple options for users.

Such an assessment is contingent on understanding a place and time. These rules did not extend to diffusion of broadband over cable and have led to a mixed experience in the diffusion of DSL as well – though, as noted, it is doubtful whether the urban biases of these technologies would differ under any set of rules. In addition, if Internet telephony becomes viable, then the present set of rules are potentially quite disruptive for the US system for assuring universal service in high cost rural areas. Certainly a revision of the revenue base for subsidizing rural telephony would have large consequences for diffusion of related information infrastructure, such as the Internet.

The other set of important rules became embedded in the Telecom Act of 1996. This Act set up the administration of the E-rate program, whose affect was unambiguously expansionary in rural areas and inner-city areas. Beyond this, the contribution of this Act is much harder to assess, at least in terms of its short run geographic impact. To be sure, it is clear that the Telecom Act of 1996, as well as the FCC's and courts interpretation of that Act, had consequences for the experience of CLECs. These decisions shaped the prices paid by CLECS, their ability to offer services such as DSL, and their ability to resolve disputes with incumbent local exchange carriers. But even if the Act had been written and interpreted differently, the particular fortunes of firms might have changed, but not necessarily the geographic dispersion of Internet infrastructure. Hence, this question awaits further analysis.

7.6. Are There Lessons for Other Countries?

Factors unique to the United States shaped the economic geography of the Internet experience in the United States: And, yet, despite these unique features, there are lessons for development of Internet infrastructure in other countries. What was unique about the US experience? As noted, many regulatory issues were first addressed in the US and were addressed in idiosyncratic ways.

In spite of this idiosyncrasy, there are three overriding lessons that have potential to move across international boundaries: First when uncertainty is irreducible, it is better to rely on private incentives to develop mass market services at a local level. Once the technology was commercialized, private firms tailored it in multiple locations in ways that nobody foresaw. Indeed, the eventual shape, speed, growth, and use of the commercial Internet was not foreseen within government circles (at NSF), despite (comparatively) good intentions and benign motives on the part of government overseers, and despite advise from the best technical experts in the world. If markets are better in this set of circumstances in the US, then a similar broad lesson probably applies elsewhere, even with all the differences found elsewhere. This principle holds with particular force for business infrastructure, where opportunities are still being developed.

Second, Internet infrastructure grew because it is malleable, not because it was technically perfect. In many places it is better thought of as a cheap retrofit on top of the existing communications infrastructure. No single solution was right for every situation, but a TCP/IP solution could be found in most places. Hence, the technology appealed to a wide variety of potential users, significantly helping the technology transform from being a niche use into mass-market applications. Stated succinctly, malleability was important because demand-side economies of scale could not be achieved with a perfect technical fix; it could only be achieved with a technology that could be deployed in multiple locations.

Third, there is no optimal combination of unfettered behavior and regulated rules for encouraging Internet growth, but exclusive use of either markets or government will surely fail in most countries. In the US it would not have been possible for any single administrative agency to build and manage such a network under government auspices. Indeed, one of the smartest things the NSF did was decide to give up managing the whole infrastructure, while putting in place a few scalable features, such as an address system and data-exchange points. Similarly, an unfettered system simply could not have worked without government help. For example, the competitiveness of the backbone network within the US almost surely arose due to the novelty of the market and the sheer size of demand. This will not be replicated in many small countries, which will have sufficient demand to support only a small number of backbone providers. In that case, government regulation has a role to play in stopping anti-competitive behavior in interconnection practices.

References

Abel, Jaison and Matthew Clements, 2001, "Entry Under Asymmetric Regulation," *Review of Industrial Organization*, 19(2), 227-242.

Augereau, A.; Greenstein, S., 2001, "The Need for Speed in Emerging Communications Markets: Upgrades to Advanced Technology at Internet Service Providers." *International Journal of Industrial Organization*, 19, 1085-1102.

Bailey, Joseph, and Lee McKnight, 1997, "Scalable Internet Interconnection Agreements and Integrated Services," in (eds) Brian Kahin and James H. Keller, *Coordinating the Internet*, MIT Press.

Beardsell, M.; Henderson V., 1999, "Spatial Evolution of the Computer Industry in the USA." *European Economic Review*, 43(2), pp. 431-456.

Bertot, John, and Charles McClure, 2000, *Public Libraries and the Internet, 2000,* Reports prepared for National Commission on Libraries and Information Science, Washington DC., http://www.nclis.gov/statsurv/2000plo.pdf.

Besen Stanley, Paul Milgrom, Bridger Mitchell, and Padmanahaban Sringagesh, 2001, "Advances in Routing Technologies and Internet Peering Agreements," *American Economic Review,* May, pp 292-296.

Besen, Stanley, Jeffrey S. Spigel, and Padmanabhan Srinagesh, 2002, "Evaluating the Competitive Effects of Mergers of Internet Backbone Providers," *ACM Transactions on Internet Technology*

Blumenthal, Marjory S., and David D. Clark, 2001, "Rethinking the Design of the Internet: The End-to-End Arguments vs. The Brave New World." In (eds.) Benjamin Compaine and Shane Greenstein, *Communications Policy in Transition: The Internet and Beyond.* MIT Press: Cambridge, MA.

Boardwatch Magazine, various, *Directory of Internet Service Providers*, Littleton, CO.

Boczkowski, Pablo, 2002, "The Development and Use of Online Newspapers: What Research Tells Us and What We Might Want to Know," in (eds) Leah Lievrouw and Sonia Livingstone, *The Handbook of New Media: Social Shaping and Consequences of ICTs*, Sage, London, pp 270-286.

Boczkowski, Pablo, 2003, *Emerging Media: Innovation in Online Newspapers*, MIT Press, Cambridge, MA.

Bresnahan, Timothy, Erik Brynjolfsson, and Lorin Hitt. 2002. Information Technology, Work Organization, and the Demand for Skilled Labor: Firm-Level Evidence. *Quarterly Journal of Economics* 117 (February): 339-376.

Bresnahan, Timothy, and Shane Greenstein, 2001, "The Economic Contribution of Information Technology: Towards Comparative and User Studies." *Evolutionary Economics*, 11, pp. 95-118.

Bresnahan, Timothy, and Manuel Trajtenberg. 1995. General Purpose Technologies: 'Engines of Growth'? *Journal of Econometrics.* 65 (1): 83-108.

Buckley, Patricia and Sabrina Montes, 2002, *Main Street in the Digital Age: How Small and Medium Sized Businesses are Using the Tools of the New Economy,* US Department of Commerce, Economics and Statistics Administration, Washington, DC.

Cairncross, Frances. 1997. *The Death of Distance.* Cambridge, MA: Harvard Press.

Cannon, Robert, 2001, "Where Internet Service Providers and Telephone Companies Compete: A Guide to the Computer Inquiries, Enhanced Service Providers, and Information Service Providers," In (eds) Benjamin Compaine and Shane Greenstein, *Communications Policy in Transition: The Internet and Beyond.* MIT Press: Cambridge.

Carr, Nicholas, 2000, "An Interview with Akamai's George Comrades," *Harvard Business Review* (May-June), pp. 118-125.

Castells, Manuel. 2002. *The Internet Galaxy, Reflections on the Internet, Business and Society*, Oxford University Press.

Cawley, Richard A., 1997, "Interconnection, Pricing, and Settlements: Some Healthy Jostling in the Growth of the Internet," in (eds) Brian Kahin and James H. Keller, *Coordinating the Internet*, MIT Press.

Chapple, K.; Zook, M.A., 2002, "Why Some IT Jobs Stay: The rise of Job Training in Information Technology." *Journal of Urban Technology*, 9(1), pp. 57-83.

Cherry, Barbara, Alan H. Hammond, and Steve Wildman, 1999. *Making Universal Service Policy: Enhancing the Process Through Multidisciplinary Evaluation*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chinoy, Bilal, and Timothy J. Salo, 1997, "Internet Exchanges: Policy-Driven Evolution," in (eds) Brian Kahin and James H. Keller, *Coordinating the Internet*, MIT Press.

Chyi, H.I.; Sylvie, G. 1998, "Competing with Whom? And How? A Structural Analysis of the Electronic Newspaper Market," *The Journal of Media Economics, 11(2), 1-18.*

Chyi, H.I.; Sylvie, G., 2001, "The Medium is Global, the Content is Not: The Role of Geography in Online Newspaper Markets." *Journal of Media Economics*, 14(4), pp. 231-248.

Claffy, K. 2001, "CAIDA: Visualizing the Internet," http://www.caida.org/outreach/papers/2001/caida/

Claffy, K, 2000, "Measuring the Internet," *IEEE Internet Computing,* http://www.caida.org/outreach/papers/2000/ieee0001/

Coffman, K. G., and Andrew Odlyzko, 2002, "Growth of the Internet," in (eds) I.P Kaminow and T. Li, *Optical Fiber Telecommunications IV B: Systems and Impairments,* Academic Press, pp 17-56.

Compaine, Benjamin. 2001. *The Digital Divide: Facing a Crisis or Creating a Myth?* Cambridge, MA: MIT Press.

Cortada, James W., 1996, *Information Technology as Business History: Issues in the History and Management of Computers.* Westport, CT: Greenwood Press.

Cortada, James, W., 2004, *The Digital Hand: How Computers Changed the Work of American Manufacturing, Transportation and Retail Industries.* Oxford University Press, New York.

Crandall, Roberts W., 2001, "An Assessment of the Competitive Local Exchange Carriers five Years After the Passage of the Telecommunications Act" A Report for the USTA by Criterion Economics, LLC.

Crandall, Robert W. and James H. Alleman, 2002, *Broadband: Should we Regulate High-Speed Internet Access?* AEI-Brookings Joint Center for Regulatory Studies, Washington DC

Crandall, Robert W. and J. Gregory Sidak, 2002, "Is Structural Separation of Incumbent Local Exchange Carriers Necessary for Competition?" *Yale Journal on Regulation*, 19(2), 1-75.

Crandall, Robert W. and Leonard Waverman, 2000, *Who Pays for Universal Service? When Telephone Subsidies Become Transparent*, Brookings Institution Press, Washington DC

David, Paul, and Shane Greenstein, 1990, "The Economics of Compatibility of Standards: A Survey." *Economics of Innovation and New Technology*, Vol. 1., pp 3-41.

Dodge, Martin, 2001, "Cybergeography." *Environment and Planning B-Planning & Design*, 28(1), pp. 1-2.

Dodge, Martin and Rob Kitchin, 2001a, *Atlas of CyberSpace,* Addison-Wesley, London.

Dodge, Martin and Rob Kitchin, 2001b, *Mapping CyberSpace,* Routledge, London.

Downes, Tom, and Greenstein, Shane, 1998, "Do Commercial ISPs Provide Universal Access?" in (Ed) Sharon Gillett and Ingo Vogelsang, *Competition, Regulation and Convergence: Current Trends in Telecommunications Policy Research,* Lawrence Erlbaum Associates, pp. 195-212.

Downes, T.; and Shane Greenstein, 2002, "Universal Access and Local Internet Markets in the U.S." *Research Policy*, 31, pp. 1035-1052.

Dumais, G.; Ellison, G.; Glaeser, E.L., 2002, "Geographic Concentration as a Dynamic Process." *Review of Economics and Statistics*, 84(2), pp. 193-204.

James Eisner, Tracy Waldon, 2001, "The demand for bandwidth: second telephone lines and on-line services." *Information Economics and Policy* 13 (3): 301-309.

Esbin, Barbara, 1998, "Internet Over Cable: Defining the Future in Terms of the Past," Federal Communications Commission, Office of Planning and Policy, Working Paper No. 30,

Farnon, Maria, and Scott Huddle, 1997, "Settlement Systems for the Internet," in (eds) Brian Kahin and James H. Keller, *Coordinating the Internet*, MIT Press.

Faulhaber, Gerald, and Christiaan Hogendorn, 2004, "The Market Structure of Broadband Telecommunications," *Journal of Industrial Economics*, V 48 (3), pp. 305-330.

Federal Communication Commission, 2000, "Local Telephone Competition: Status as of June 30, 2000," Industry Analysis Division, Common Carrier Bureau, available at http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/recent.html

Federal Communication Commission, 2001, "Local Telephone Competition: Status as of June 30, 2001," Industry Analysis Division, Common Carrier Bureau, available at http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/recent.html

Feldman, M.P., 2002, "The Internet Revolution and the Geography of Innovation." *International Social Science Journal*, 54(1).

Fitoussi, David, 2004, "So Close and Yet So Far: Information Technology and the Spatial Distribution of Work," Mimeo, MIT.

Forman, Chris, 2002. The Corporate Digital Divide: Determinants of Internet Adoption. Working Paper, Graduate School of Industrial Administration, Carnegie Mellon University. Available at http://www.andrew.cmu.edu/~cforman/research/corp_divide.pdf.

Forman, Chris, Avi Goldfarb, and Shane Greenstein. 2003a. The Geographic Dispersion of Commercial Internet Use. In *Rethinking Rights and Regulations: Institutional Responses to New Communication Technologies*, ed. Steve Wildman and Lorrie Cranor, forthcoming, Cambridge: MIT Press.

---, 2003b, "Which Industries use the Internet?" in (ed) Michael Baye, *Organizing the New Industrial Economy,* Elsevier. Pages 47-72.

---, 2003c, "How did Location Affect adoption of the Internet by Commercial Establishments? Urban density, Global Village and Industry Composition". NBER Working Paper # 9979. Cambridge, MA.

Frazer, Karen, 1995, *NSFNET: A Partnership for High-Speed Networking, Final Report, 1987-1995.* Merit Networks.

Frieden, Rob, 2001, "The Potential for Scrutiny of Internet Peering Policies in Multilateral Forums," in (eds) Benjamin Compaine and Shane Greenstein, *Communications Policy in Transition: The Internet and Beyond.* MIT Press: Cambridge.

Gabel, D.; Kwan, F., 2001,"Accessibility of Broadband Communication Services by Various Segments of the American Population." In (eds) (eds) Benjamin Compaine and Shane Greenstein, *Communications Policy in Transition: The Internet and Beyond.* MIT Press: Cambridge, MA.

Garcia, D. L., 1996, "Who? What? Where? A Look at Internet Deployment in Rural America," Rural Telecommunications, November/December, 25-29.

Gillett, S.E., 2000, "Universal Service: Defining the Policy Goal in the Age of the Internet." *Information Society*, 16(2), pp. 147-149.

Glaeser, E., and Gaspar, J., 1997, "Information Technology and the Future of Cities." *Journal of Urban Economics*, 43, pp. 136-156.

Goldfarb, Avi, 2004. Concentration in Advertising-Supported Online Markets: An Empirical Approach. Economics of Innovation and New Technology 13(6), 581-594.

Goolsbee, Austan, and Peter Klenow, 1999, "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." NBER working paper # 7329.

Goolsbee, Austan, and Jonathan Guryan, 2002, "The Impact of Internet Subsidies for Public Schools," 2002, NBER Working Paper #9090.

Gorman, Sean P., Forthcoming, "Where are the Web Factories: The Urban bias of E-Business Location." *Tijdschrift voor Economische Geografie.*

Gorman, S.P.; Malecki, E.J., 2000, "The Networks of the Internet: an Analysis of Provider Networks in the USA." *Telecommunications Policy*, 24(2), pp. 113-134.

Gorman, Sean, Malecki, Edward J., 2002, "Fixed and Fluid: Stability and Change in the Geography of the Internet," *Telecommunications Policy*, 26, pp 389-413.

Gorman, Sean, Angela McIntee, forthcoming, "Tethered Connectivity? The Spatial Distribution of Wireless Infrastructure," *Environment and Planning A.*

Gregg, Billy Jack, 2002, "A Survey of Unbundled Network Element Prices in the United States," Public Service Commission of West Virginia.

Greenstein, Shane, 2000a. "Building and Delivering the Virtual World: Commercializing Services for Internet Access." *The Journal of Industrial Economics*, 48(4), 391-411.

Greenstein, Shane, 2000b, "Empirical Evidence on Commercial Internet Access Providers' Propensity to Offer new Services," (ed.) Benjamin Compaign and Ingo Vogelsang. *The Internet Upheaval, Raising Questions and Seeking Answers in Communications Policy.* MIT Press, Cambridge.

Greenstein, S. M., 2001, "Commercialization of the Internet: The Interaction of Public Policy and Private Actions," in: A. Jaffe, J. Lerner and S. Stern (Editors), *Innovation, Policy and the Economy,* MIT Press, Cambridge.

Greenstein, Shane and Michael Mazzeo, 2003, "Differentiation Strategy and Market Deregulation: Local Telecommunication Entry in the Late 1990s," NBER Working Paper 9761, Cambridge.

Greenstein, Shane, and Mercedes Lizardo, 1999, "Determinants of the Regional

Distribution of Information Technology Infrastructure in U.S.," in (Eds) Dale Orr and Tom Wilson, *The Electronic Village: Public Policy Issues of the Information Economy.* C.D. Howe Institute, Toronto, Canada.

Greenstein, Shane, Mercedes Lizardo and Pablo Spiller, 2003, "The Evolution of Advanced Large Scale Information Infrastructure in the United States," in (ed.) Allan Shampine, *Down to the Wire: Studies in the Diffusion and Regulation of Telecommunications Technologies*, Nova Press.

Grubesic, Tony, 2003, "Spatial Dimensions of Internet Activity," *Telecommunications Policy,* 26, pp 363-387.

Grubesic, Tony, 2003, "Inequities in the Broadband Revolution," *The Annals of Regional Science,* 37, pp 263-289.

Grubesic, T.H.; Murray, A.T., 2002, "Constructing the Divide: Spatial Disparities in Broadband Access." *Papers in Regional Science*, 81(2), pp. 197-221.

Grubesic, T.H.; O'Kelly, M.E., 2002, "Using Points of Presence to Measure Accessibility to the Commercial Internet." *Professional Geographer*, 54(2), pp. 259-278.

Grubesic, Tony, Murray O'Kelly, Alan Murray, 2003, "A Geographic Perspective on Commercial Internet Survivability," *Telematics and Informatics,* 20, pp 51-69.

Hausman, Jerry A., J. Gregory Sidak, and Hal J. Singer, 2001 "Cable Modems and DSL: Broadband Internet Access for Residential Customers," in *American Economic Review,* May, pp 302-307.

Henry, David, and Donald Dalton, 2002, "Information Technology Industries in the New Economy," Chapter III in Digital Economy 2002, Department of Commerce, at http://www.esa.doc.gov/reports.cfm.

Hindman, D.B., 2000, "The Rural-Urban Digital Divide." *Journalism & Mass Communication Quarterly*, 77(3), pp. 549-560.

Hogendorn, Christiaan, 2003, "Excessive (?) Entry of National Telecom Networks, 1990-2001," mimeo, Wesleyan University.

Huffaker, B., M. Fomenkov, D. Plummer, D. Moore, and K. Claffy, 2002, *Distance Metrics in the Internet*, Presented at the IEEE International Telecommunications Symposium, 2002, copy at. http://www.caida.org/outreach/papers/2002/Distance/.

Hurley, Deborah, and James H. Keller, 1999, *The First 100 Feet: Options for Internet and Broadband Access*, MIT Press.

Jacobs, Jane, 1969, *The Economy of Cities.* New York: Vintage.

Juliussen, Egil, and Karen Petska-Juliussen, 1998, *Internet Industry Almanac*, Peer-to-peer Communications; San Jose, CA.

Kahin, Brian, 1992, *Building Information Infrastructure, Issues in the Development of the National Research and Education Network,* McGraw-Hill Primis, Cambridge, MA.

Kahin, Brian, 1997 “The U.S. National Information Infrastructure Initiative: The Market, the Web, and the Virtual Project,” in (eds) Brian Kahin and Ernest Wilson, *National Information Infrastructure Initiatives, Vision and Policy Design,* MIT Press.

Kahin, Brian and Janet Abbate, 1995, *Standards for Information Infrastructure,* MIT Press, Cambridge, MA.

Kahin, Brian, and James H. Keller, 1997, *Coordinating the Internet*, MIT Press.

Kahin, Brian, and Charles Nesson, 1999, *Borders in Cyberspace: Information Policy and the Global Information Infrastructure*, The MIT Press.

Kahn, Robert, 1995, “The Role of Government in the Evolution of the Internet,” in National Academy of Engineering, *Revolution in the U.S. Information Infrastructure*, National Academy Press, Washington, D.C.

Kende, Michael, 2000, “The Digital Handshake: Connecting Internet Backbones,” Federal Communications Commission, Office of Planning and Policy, Working Paper No. 32. Washington, D.C.

Kenney, Martin, 2003, “The Growth and Development of the Internet in the United States,” in (ed) Bruce Kogut, *The Global Internet Economy*, MIT Press, Cambridge.

Kitchin, R.M., 1998, “Towards Geographies of Cyberspace.” *Progress in Human Geography*, 22(3), pp. 385-406.

Kitchin, R., 1999, “Digital Places: Living with Geographic Information Technologies.” *Progress in Human Geography*, 23(4), pp. 647-648.

Kitchin, R.M.; Dodge, M., 2001, “’Placing’ Cyberspace: Why Geography Still Matters.” *Information Technology, Education and Society*, 1(2), pp. 25-46.

Kolko, Jed, 2000, “The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage,” in (eds) Ingo Vogelsang and Benjamin Compaine, *The Internet Upheavel: Raising Questions, Seeking Answers in Communications Policy,* MIT Press, Cambridge, MA.

Kolko, Jed, 2002. “Silicon Mountains, Silicon Molehills, Geographic Concentration and Convergence of Internet Industries in the U.S.,” *Economics of Information and Policy.*

Koski, Heli and Sumit Majumdar, 2002, “Paragons of Virture? Competitor Entry and the Strategies of Incumbents in the US Local Telecommunications Industry,” *Information Economics and Policy*, 14, pp 453-480.

Kotkin, Joel, 2000. *The New Geography: How the Digital Revolution is Reshaping the American Landscape*, New York: Random House.

Krugman, Paul, 1991, *Geography and Trade*. Cambridge: MIT Press.

Kruse, Hans, William Yurcik, and Lawrence Lessig, 2001, "The InterNAT : Policy Implciations of the Internet Architecture Debate," in (eds) Benjamin Compaine and Shane Greenstein, *Communications Policy in Transition: The Internet and Beyond.* MIT Press: Cambridge, MA.

Laffont, Jean-Jacques, and Jean Tirole, 2000. *Competition in Telecommunications*, MIT Press, Cambridge.

Laffont, Jean-Jacques, Scott Marcus, Patrick Rey, and Jean Tirole, 2003. "Internet interconnection and the off-net-cost pricing principle", *RAND Journal of Economics*, Vol. 34, No. 2.

Lehr, William, and Sharon Gillett, 2000 "Availability of Broadband Internet Access: Empirical Evidence." www.ksg.harvard.edu/iip/access/gillett_lehr.PDF.

Lemley, Mark, and Larry Lessig, 2001, "The End of End-to-End" Preserving the Architecture of the Internet in the Broadband Era," 48 UCLA Law Review, 925 - 948.

Light, J. 2001, "Rethinking the Digital Divide." *Harvard Educational Review*, 71(4), pp. 710–734.

Light, J. 1999, "From City Space to Cyberspace" in *Virtual Geographies*, Eds. Mike Crang, Philip Crang, and Jon May. London, Routledge.

Mackie-Mason, Jeff, and Hal Varian, 1995, "Pricing the Internet," in (eds) Brian Kahin and James Keller, *Public Access to the Internet,* MIT Press, Cambridge.

Malecki, Edward, J., 2002a. "The Economic Geography of the Internet's Infrastructure," *Economic Geography*, V78, 4, pp 399-424.

Malecki, Edward, J., 2002b, "Location Competition in Telecommunications in the United States: Supporting Conditions, Policies, and Impacts," *The Annals of Regional Science,* V 36, pp 437-454.

Malecki, Edward, J., 2002c, "Hard and Soft Networks for Urban Competitiveness," *Urban Studies*, 39, 5-6, pp 929-945.

Malecki, Edward, J., 2003, "Digital Development in Rural Areas: Potential and Pitfalls," *Journal of Rural Studies,* 19, pp 201-214.

Malecki, Edward, J., and Sean Gorman, 2001, "Maybe the Death of Distance, but not the End of Geography: The Internet as a Network," in (eds) Thomas Leinbach and Stanley Brunn, *Worlds of E-Commerce: Economic, Geographical and Social Dimensions*, John Wiley and Sons, LTD, New York.

Maloff Group International, Inc., 1997, *1996-1997 Internet Access Providers Marketplace Analysis*, Maloff Group International, Inc., Dexter, MO.

Meeker, Mary, and Chris DePuy, 1996, *The Internet Report*, Harper Business, New York.

Milgrom, Paul, Bridger Mitchell, and Padmanabhan Srinagesh, 2000, “Competitive Effects of Internet Peering Policies,” in (eds) Ingo Vogelsang and Benjamin Compaine, *The Internet Upheaval: Raising Questions Seeking Answers in Communications Policy,* MIT Press.

Mini, Federico, 2001, “The Role of Incentives for Opening Monopoly Markets: Comparing GTE and BOC Cooperation with Local Entrants,” *Journal of Industrial Economics*, 49(3), 379-413.

Moss, M.L. and Townsend, 1997, A. “Tracking the Net: Using Domain Names to Measure the Growth of the Internet in U.S. Cities.” *Journal of Urban Technology*, 4(3), pp. 47-60.

Moss, M.L. and Townsend, A.M., 1998, “The Role of the Real City in Cyberspace: Understanding Regional Variations in Internet Accessibility and Utilization.” Paper presented at the Project Varenius Meeting on Measuring and Representing Accessibility in the Information Age, Pacific Grove, California.

Moss, M.L. and Townsend, A.M., 1999, “How Telecommunications Systems are Transforming Urban Spaces.” in *Fractured Geographies: Cities in the Telecommunications Age*, James O. Wheeler and Yuko Aoyama, eds., Routledge, New York.

Moss, M.L. and Townsend, A.M., 2000a, “The Internet Backbone and the American Metropolis.” *Information Society*, 16(1), pp. 35-47.

Moss, M.L. and Townsend, A.M., 2000b, *The Role of the Real City in Cyberspace: Measuring and Representing Regional variations in Internet Accessibility, Information, Place, and Cyberspace*, Donald Janelle and David Hodge, eds., Springer-Vedag, Berlin, 2000.

Mowery, D.C., and T.S. Simcoe, 2002, "The Origins and Evolution of the Internet," in (eds) B. Steil, D. Victor, and R. Nelson, *Technological Innovation and Economic Performance:* Princeton University Press, pp. 229-264.

National Research Council, 2001, *The Internet’s Coming of Age*, National Academy Press, Washington D.C.

National Research Council, 2002, *Broadband, Bringing Home the Bits,* National Academy Press, Washington, D.C.

New Paradigm Resources Group, 2000, *CLEC Report*, Chicago, IL.

Nicholas, Kyle. H., 2000, “Stronger then Barbed Wire: How Geo-Policy Barriers Coinstruct Rural Internet Access,” in (eds) Lorrie Faith Cranor and Shane Greenstein, *Communications Policy and Information Technology: Promises, Problems, Prospects,* MIT Press, Cambridge, MA.

Noam, Eli, 2001, *Interconnecting the Network of Networks,* MIT Press.

Noam, Eli, 2002, “Interconnection Practices,” in (eds) Martin Cave, Sumit Majumdar, Ingo Vogelsang, *Handbook of Telecommunications Economics, Volume 1*, North-Holland, Amsterdam.

Noll, Roger, Dina Older-Aguilar, Gregory Rosston, and Richard Ross. 2001. "The Digital Divide: Definitions, Measurement, and Policy Issues," Paper presented at the American Economics Association Meetings, January 2002.

National Telecommunications and Information Administration (NTIA), 1995. “Falling Through the Net: A Survey of the ‘Have Nots’ in Rural and Urban America.” http://www.ntia.doc.gov/reports.html.

--, 1997, “Falling Through the Net: Defining the Digital Divide.” http://www.ntia.doc.gov/reports.html.

--, 1998. “Falling Through the Net II: New Data on the Digital Divide.” http://www.ntia.doc.gov/reports.html.

--, 2000. “Advanced Telecommunications in Rural America, The Challenge of Bringing Broadband Service to All Americans.” See http://www.ntia.doc.gov/reports.html.

--, 2002. “A Nation Online: How Americans Are Expanding Their Use Of The Internet.” http://www.ntia.doc.gov/reports.html.

Odlyzko, Andrew, 2001, “Internet growth: Myth and reality, use and abuse,” *J. Computer Resource Management*, 102, Spring. Pp. 23-27.

O’Kelly, Morton, and Tony Grubesic, 2002, “Backbone Topology, Access, and the Commercial Internet, 1997-2000,” *Environment and Planning B: Planning and Design,*, V 29, pp. 533-552.

Oxman, Jason, 1999, “The FCC and the Unregulation of the Internet,” Federal Communications Commission, Office of Planning and Policy, Working paper 31.

Parker, E.B., 2000, “Closing the Digital Divide in Rural America.” *Telecommunications Policy*, 24(4), pp. 281-290.

Premkumar, G.; Roberts, M., 1999, “Adoption of New Information Technologies in Rural Small Businesses.” *Omega-International Journal of Management Science*, 27(4), pp. 467-484.

Price, Lee, with George McKittrick, 2002, “Setting the Stage: The New Economy Endures Despite Reduced IT Investment,” in Chapter I, Digital Economy 2002, Department of Commerce, at http://www.esa.doc.gov/reports.cfm.

Prieger, James E., 2001, “The Supply Side of the Digital Divide: Is There Redlining in the Broadband Internet Access Market?” AEI-Brookings Joint Center for Regulatory Studies, Working Paper 01-16, December 2001.

Rogers, Everett M. 1995. *Diffusion of Innovations.* New York: Free Press.

Rosston, Greg, and Brad Wimmer, 2001, "'From C to Shining C' Competition and Cross-Subsidy in Communications," in (eds) Benjamin Compaine and Shane Greenstein, *Communications Policy in Transition: The Internet and Beyond*, MIT Press.

Shampine, Allan, 2001, "Determinants of the Diffusion of US Digital Telecommunications," *Journal of Evolutionary Economics*, V11, pp. 249-261.

Shiman, Daniel, and Jessica Rosenworcel, 2002, "Assessing the Effectiveness of Section 271 Five Years After the Telecommunications Act of 1996", in (eds) Lorrie Cranor and Shane Greenstein, *Communications Policy and Information Technology: Promises, Problems and Prospects*, MIT Press.

Sidak Gregory, 2003, "The Failure of Good Intentions: The WorldCom Fraud and the Collapse of American Telecommunications After Deregulation," *Yale Journal of Regulation*, V20, pp 207-267.

Sidak, Gregory and Dan Spulber, 1998, "Cyberjam: The Law and Economics of Internet Congestion of the Telephone Network." *Harvard Journal of Law and Public Policy*, 21, 2.

Sinai, Todd and Joel Waldfogel, 2000, "Geography and the Internet: Is the Internet a Substitute or a Complement for Cities?" Working Paper, Wharton School of Business, University of Pennsylvania, Philadelphia.

Spulber, Daniel, and Christopher Yoo, 2003, "Access to Networks: Economic and Constitutional Connections," *Cornell Law Review*, 2003, v. 88.

Srinagesh, P., 1997, Internet Cost Structures and Interconnections Agreements, in: L.W. McKnight and J. Bailey (Editors), *Internet Economics*, MIT Press, Cambridge, MA.

Strover, Sharon, 2001, "Rural Internet Connectivity." *Telecommunications Policy*, 25(5), pp. 331-347.

Strover, Sharon, and Lon Berquist, "Developing Telecommunications Infrastructure: State and Local Policy Collisions," in (eds)) Benjamin Compaine and Shane Greenstein, *Communications Policy in Transition: The Internet and Beyond,* MIT Press.

Strover, Sharon, Michael Oden and Nobuya Inagaki, 2002, Telecommunications and Rural Economies: Findings from the Appalachian Region, (eds) Lorrie Faith Cranor and Shane Greenstein, *Communication Policy and Information Technology: Promises, Problems, Prospects.* MIT Press, Cambridge, MA.

Suhonen, Susanne, 2002, *Industry Evolution and Shakeout Mechanisms: The Case of the Internet Service Provider Industry,* Dissertation, Helsinki School of Economics.

Townsend, A.M., 2001a, "Network Cities and the Global Structure of the Internet." *American Behavioral Scientist*, 44(10), pp. 1697-1716.

Townsend, A.M., 2001b, "The Internet and the Rise of the New Network Cities, 1969-1999." *Environment and Planning B-Planning & Design*, 28(1), pp. 39-58.

U.S. Department of Agriculture, 2000. *Advanced Telecommunications in Rural America, The Challenge of Bringing Broadband Communications to All of America,* http://www.ntia.doc.gov/reports/ruralbb42600.pdf

Waldrop, M. Mitchell, 2002, *The Dream Machine, J.C.R. Licklider and the Revolution that Made Computing Personal*, Penguin Books.

Warf, B., 2001, "Segueways into Cyberspace: Multiple Geographies of the Digital Divide." *Environment and Planning B-Planning & Design*, 28(1), pp. 3-19.

Weinberg, Jonathan, 1999. "The Internet and Telecommunications Services, Access Charges, Universal Service Mechanisms, and Other Flotsam of the Regulatory System." V16, *Yale Journal of Regulation*, Spring.

Weinberg, Jonathan, 2001, "Internet Telephony Regulation," in (eds) Lee McKnight, William Lehr, and David Clark, *Internet Telephony*, MIT Press.

Werbach, Kevin, 1997. *A Digital Tornado: The Internet and Telecommunications Policy.* FCC, Office of Planning and Policy Working Paper 29, March.

Wheeler, D.C.; O'Kelly, M.E., 1999, "Network Topology and City Accessibility of the Commercial Internet." *Professional Geographer*, 51(3), pp. 327-339.

Woroch, Glenn, 2001, "Local Network Competition," in (eds) Martin Cave, Sumit Majumdar, and Ingo Vogelsang*, Handbook of Telecommunications Economics*, Elsevier Publishing.

Zolnierek, James, James Eisner and Eller Burton, 2001, "An Empirical Examination of Entry Patterns in Local Telephone Markets*," Journal of Regulatory Economics*, 19(2), 143-159.

Zook, M.A., 1998, "The Web of Consumption: The Spatial Organization of the Internet Industry in the United States." Paper presented at the Association of Collegiate Schools of Planning 1998 Conference, Pasadena, CA.

Zook, M.A., 2000, "The Web of Production: the Economic Geography of Commercial Internet Content Production in the United States." *Environment and Planning A*, 32(3), pp. 411-426.

Zook, M.A., 2001, "Old Hierarchies or New Networks of Centrality? The Global Geography of the Internet Content Market." *American Behavioral Scientist*, 44(10), pp. 1679-1696.

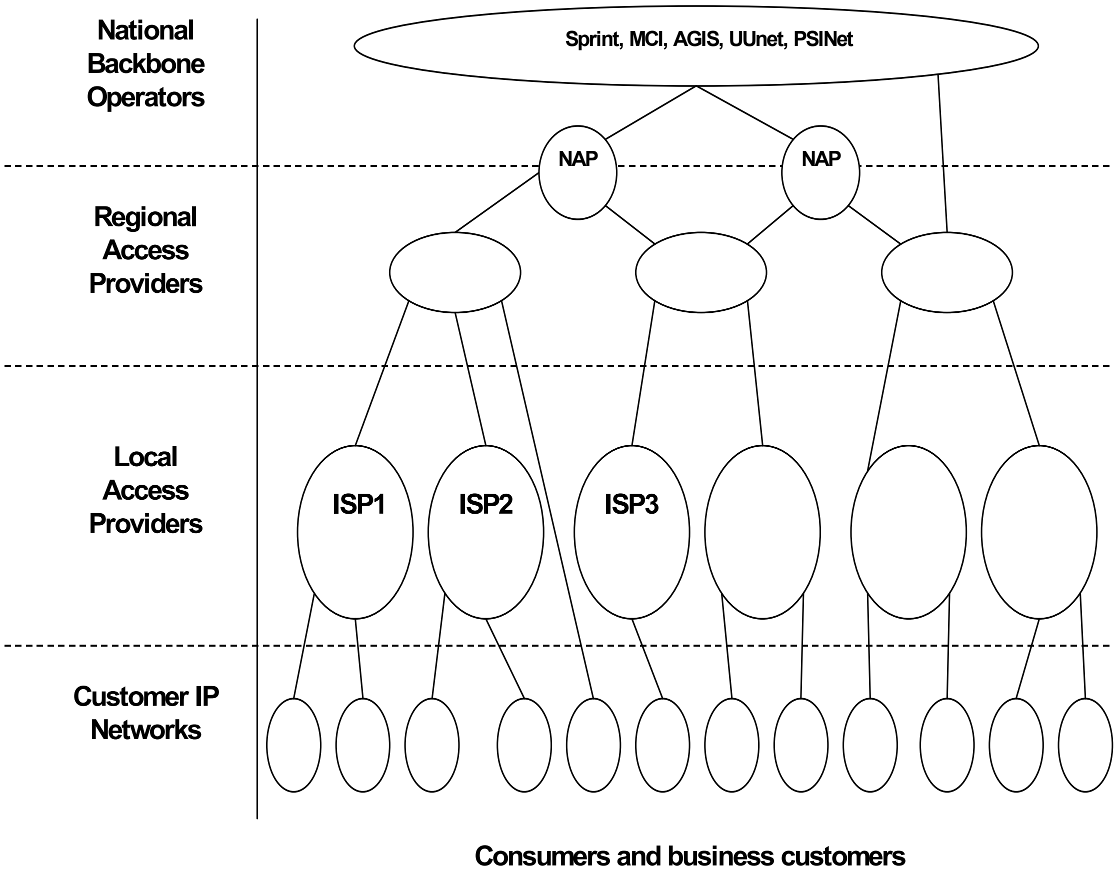
Zook, M.A., 2002, "Being Connected is a Matter of Geography." *netWorker*, 5(3), pp. 13-16.

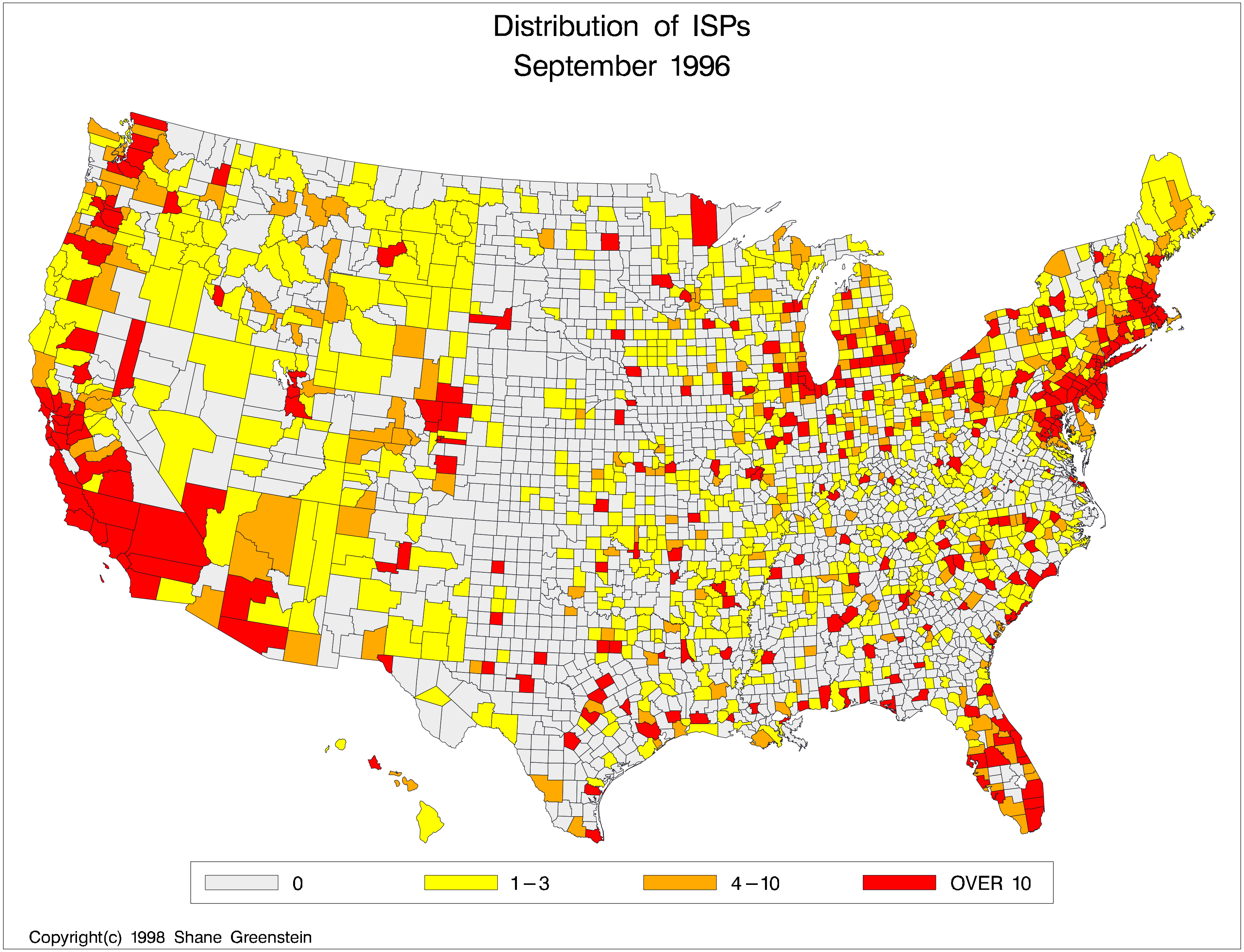
Table 1: Abridged Chronology of the first decade of the Commercial Internet

1988 NSF bids out contracts to build the NSFNet as a backbone for researchers.
1989 Internet host computers reaches 100K
1990 Arpanet decommissioned, leaving a network with over 300K host computers.
1991 NSF begins commercialization by lifting restrictions on commercial use
1992 Internet host computers reach 1 million
1992 First commercial ISPs begin to interconnect legally with Internet.
1993 NSCA at University of Illinois releases the browser named Mosaic
1994 April, Netscape is founded
1994 May, first World Wide Web Consortium (W3C) conference held
1994 August, Compuserve offer Internet access to their users
1994 Internet host computers pass 4 million
1994 November, MCI introduces an Internet connection service, AOL buys ANS
1995 February, Yahoo founded as a commercial service
1995 March, BBN rolls out first nationwide Internet service – with over 500 access points
1995 May, Uunet and PSI go public.
1995 NSFNet ends contract with ANS, MCI becomes largest Internet backbone provider
1995 August, Netscape goes public, valued at 2.2Billion on first day of trading.
1995 September, AT&T launches business Internet services.
1995 October, InternNIC begins charging for domain name registrations
1996 Telecom Act passes US Congress.
1996 January, 100K web sites on line
1996 February, AT&T begins to sell $20 home Internet access
1996 April, Yahoo goes public, reaching market value of $848M on first day of trading.
1996 May, MFS communications buys Uunet technologies for 2 Billion.
1996 Internet host computers passes 15M
1996 August, WorldCom offers $14.4B to buy MFS.
1996 November, AOL adopts flat rate fee of $19.95 per month for unlimited access.
1997 March, Network Solutions (NSI) registers the 1 millionth Internet domain name.
1997 May, GTE buys BBN and Genuity,
1997 May, Uunet announces it will charge small ISPs for interconnection
1997 September, WorldCom buys Compuserve's infrastructure, AOL buys its customer base
1997 October, WorldCom bids for MCI. WorldCom wins in November.
1997 Internet host computers passes 25M.
1998 January, AT&T buys Teleport.
1998 February, ITU announces 56K modem standards, ending war over specification.
1998 May, Network Solutions registers 2 Millionth domain name
1998 June, AT&T buys TCI, a cable company.
1998 Over 26% of US households have Internet access
1999 March, 5 million web sites on line.
1999 May, AT&T announces agreement to buy Media One, a cable company.
1999 Peak of the dot-com boom, over 490 IPOs take place during the year.
2000 January, AOL announces merger with Time Warner
2000 March, Nasdaq peaks above 5000, height of "dot-com bubble."
2000 April, "Dot-com crash." Many Internet IPOs withdrawn or delayed indefinitely.
2000 September, 25M web sites on line.
2000 Over 41% of US households have Internet access
2001 Internet host computers passes 100 M hosts
2001 July, AT&T receives bid from ComCast for cable assets.
2001 Estimates for retail e-commerce sales at $25 Billion in US
2001 Over 50% of US households have Internet access, less than a fifth of those are broadband
2002 July, MCI/WorldCom declares bankruptcy

Sources: Juliussen and Petska-Juliussen, 1998, Chapter 7, NTIA, 2002, E-Stats, www.census.gov/eos/www/ebusiness614.htm.

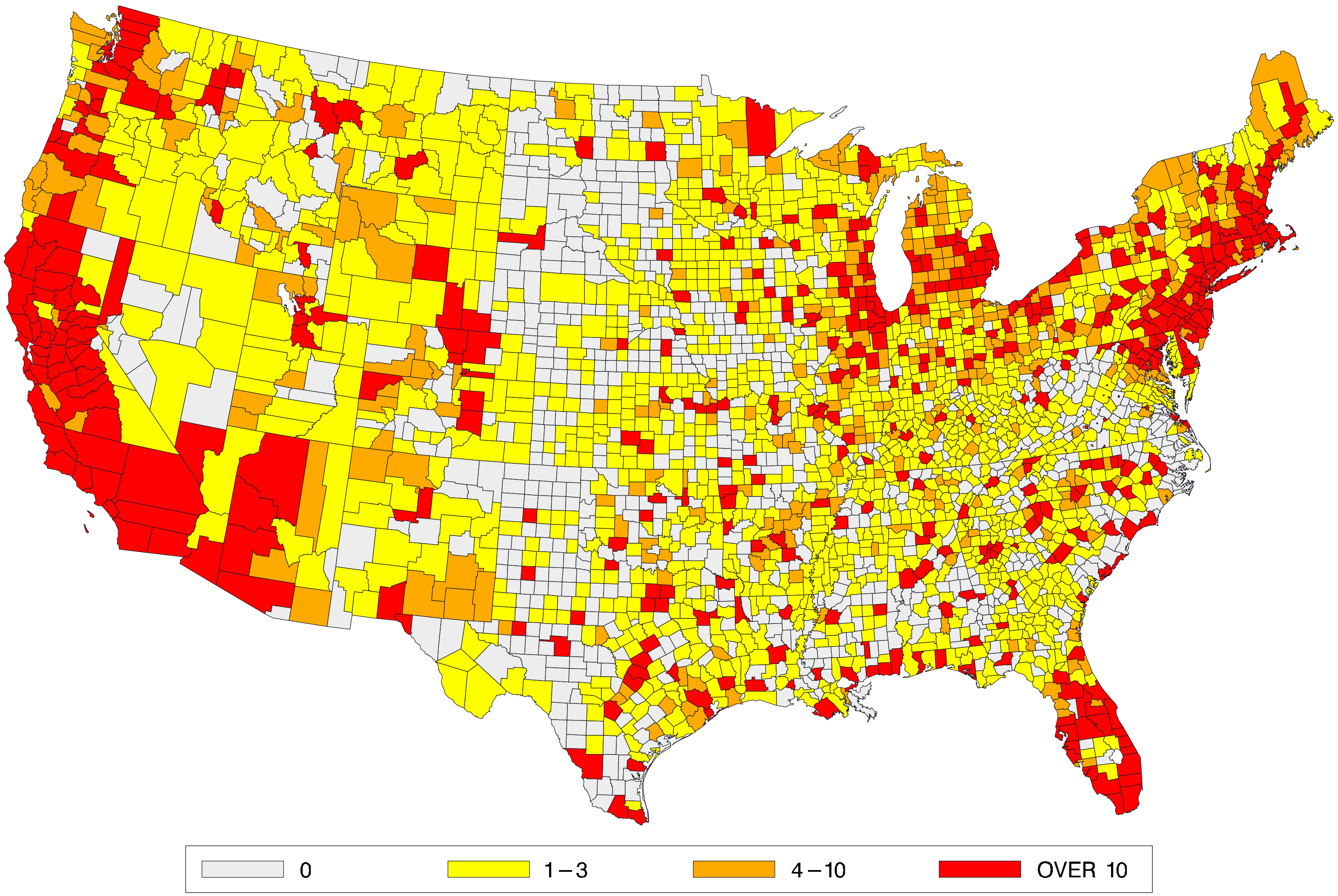
Figure 1


National
Backbone
Operators
Sprint, MCI, AGIS, UUnet, PSINet
NAP
NAP
Regional
Access
Providers
Local
Access
Providers
ISP1
ISP2
ISP3
Customer IP
Networks
Consumers and business customers



Copyright(c) 1998 Shane Greenstein

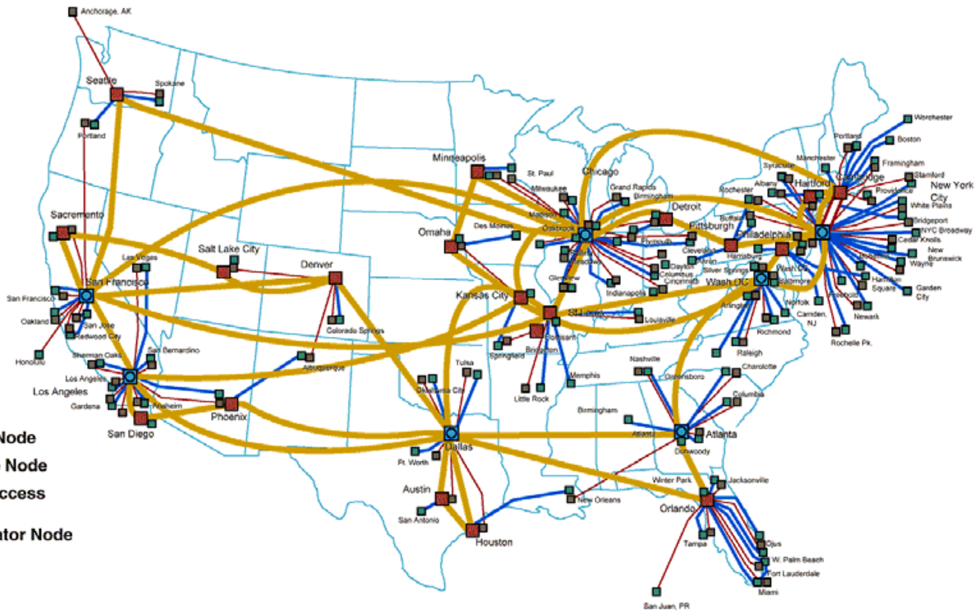
Distribution of ISPs
October 1998



Copyright(c) 1998 Tom Downes and Shane Greenstein

AT&T IP BACKBONE NETWORK

2Q2000



- Gateway Node
- Backbone Node
- Remote Access Router
- Concentrator Node (C9000)
- N X DS3
- N X OC3
- N OC48
- N X OC192

Note: map is not to scale.

UUNET's North America Internet network