Protocol Interfaces are the new bottlenecks: What the Internet means for telecom regulation

Timothy Denton, tmdenton.com
37 Heney Street, Ottawa, Ontario, Canada K1N 5V6

1. Overview

This article makes two proposals: first, that in an Internet-era, regulators have fresh and compelling duties to constrain monopoly power, which will be exercized through protocol interfaces. The second is that, owing to the separation of the applications from the transport, which is a fundamental feature of the Internet, the bit transport business will be starved for capital, and that new forms of public and private ownership of bit transport are required to solve the problem that the Internet creates. The alternative is re-monopolization, which will abort the massive creation of new wealth and the destruction of old wealth which the Internet has made possible.

This article proposes that the traditional duties of the regulator and of policy-makers, which are to limit the power of monopolies, continue in a new form, which is the supervision of the anti-competitive propensities in software architectures employed by incumbent cable and telephone local monopolies. In a software-driven system, the conceptual architecture is in embedded in code. Code determines everything important about modern communications systems, including whether they will platforms for spontaneous innovation or for the extension of monopoly power.

The second outcome which we see as inevitable is for communities to develop broadband infrastructures organized on the principles of the Internet. Just as highways are owned and managed differently from private vehicles, so will ownership and management of the transport layer be different from the ownership and management of the applications layer. These broadband infrastructures may be publicly or privately owned. They may be maintained by any form of organization suitable to the task. What will make them truly distinct from previous cable and telephone networks, however, will be the separation of the transport functions from the applications provider functions. The owner or owners of the bandwidth will not tell the users of the bandwidth what they can do with it. People will light their fiber and connect how and with whom they want, on protocols of their choosing. No one will tell the user of the network what uses the network may be put to, beyond generally applicable laws such as govern publishing.

2. Facilities-Based Competition

The policy which has been followed since the mid-1980s, and which is spreading over the world, is to encourage long distance competition in wireline telephony, cellular competition, and local competition. The latter was to be realized by the resale of network elements of the incumbent former monopoly, or facilities-based local competition, whether based in microwave or wire.

Facilities-based competition lies at the very heart of CRTC Telecom Decision 97-8, which introduced local competition in telephony in Canada.

The following quote is from this very decision, at paragraph 73:

> *The Commission is of the view that efficient and effective competition will be best achieved through facilities-based competitive service providers; otherwise, competition will only develop at the retail level, with the ILECs retaining monopoly control of wholesale level distribution.*[1]

By the end of 2001 it has become apparent, by the number of competitive local carriers going bankrupt, that facilities-based local competition is a failed policy. It has failed for many reasons. At the time it was conceived, competition was thought of in terms of offering telephone services, rather than a completely new platform for Internet-based architectures. It has also failed because of the enormous investments that must be made. Above all, it has failed because, in an Internet era, it is an irrelevant policy to the direction technology has taken. It is this last reason, in my submission, which is supreme.

In an Internet-based architecture, applications, for which people are prepared to pay, are separated from bit transport, for which, on the whole, they are not prepared to pay enough to attract investment. The separation of applications from transport is at the root of the problem of investment in local facilities. It is the hidden factor that friends of the Internet need to account for. The enemies of the Internet already understand this instinctively.

3. The end-to-end principle

The Internet grew out of a computer-communications world of academic researchers, whose assumptions about what a communications could do were shaped by the possibilities of computers[2], which are marked by exponentially declining costs of computation, storage, and – where not under monopoly control – transmission.

The Internet was built on entirely different technical assumptions than either of the other two legacy systems, circuit-switched telephony or cable television. These technical differences form the basis of the incompatibility between the wealth-creation opportunities afforded by each system. These technical differences have permitted the Internet to run on what is called "the end-to-end principle", which results in stripping out of the lower, transport-related functions of the Internet bottlenecks or control points that would determine what services or applications could be provided and sold.

The features whereby the Internet differs from telephony include:

1. The system assumes that computation, storage and transport are subject to exponentially decreasing costs, and wastes them according to their declining costs;
2. The location of intelligence is pushed as far as possible towards the edge of the network, and is owned and controlled by users to the greatest extent possible;[3]
3. The system, which is a set of instructions for machines, is layered into what I term transport, code, and content layers[4]. The lower layers, which concern themselves with the transport of signals, make the fewest possible assumptions[5] about the

nature of the task to be accomplished by the system. Hence it is optimized for no particular use, and open to the maximum feasible number of uses;

4. The relatively open content layer is a platform for innovation[6]: the World Wide Web, email, and a host of computer applications we all use testify to the transformative power of innovation without permission;
5. Because lower layers were designed to be as ignorant of the system as was consistent with the movement of traffic, the system contains very little knowledge of itself, and in particular, was not designed for record keeping necessary for billing.[7]

This is precisely the opposite doctrine to how telephone and cable systems are run, where the owners of networks determine what services are offered. In the Internet, no one has control over what services are or can be offered, at least at the technical level. of whether packets can be read, controlled, and categorized for content by the routers that guide packets to their destinations.

Numerous attacks have been made on this principle. They consist primarily of the claims of intellectual property lawyers to protect and enhance the value of copyright against the tremendous innovative force of new computer architectures, as evidenced by the attacks on file sharing programs such as Napster and Gnutella. There are also attacks on the end-to-end principle that come from network designers who seek to optimize the network for today's applications, at the expense of its general usefulness for al future applications. As David Reed writes

> “In addition to economic friction against innovation, we are creating points of control, where a new class of "trolls" are being permitted to set up shop under our network bridges. These trolls (the companies who develop, and their customers who deploy and operate these special mechanisms) must be consulted and are required to bless any new protocols or applications. Just ask a company like RealNetworks, which must negotiate with firewall vendors, ISPs and other troll-like intermediaries to clear paths for its innovative streaming media protocols. In the Internet's end-to-end design, the default situation is that a new service among willing endpoints does not require permission for deployment. But in many areas of the Internet, new chokepoints are being deployed so that anything new not explicitly permitted in advance is systematically blocked.”[8]

The restrictions on the end-to-end principle constitute one form in which the innovative possibilities of the Internet may be lessened and controlled. There exist a far more obvious source of control, however, in the service models of the telephone and cable television industries. While they were each designed for different purposes, cable and telephony share a common assumption, that the network creates the value. Services are designed and created by the network, not by the end-users. The network defines the services that may be extracted from it. Hence telephony has not acted as a platform for innovation.[9]

The following sets forth the principal differences in the two systems:

Internet

•Underspecified
•Peer-to-peer
•End-to-end
•Open
•Services are defined by anyone with an idea.

•Applications split from transport – hence the *protocol stack*

Legacy networks

•Completely specified
•Master-slave
•Service control points

•Proprietary
•Services are defined by owners of the system.
•Services are vertically integrated with transport

The telephone companies, together with their cousins in the cable television business, are masters of regulatory gaming[10]. The enormous technical and business challenge of the Internet is held at bay by the inadequacy of the means whereby we reach it, which are the two legacy networks, where abundant opportunities exist to exercise new forms of market power.

Facilities-based competition was intended to be the solution to market power exercised by the two incumbent signal delivery mechanisms. It was the means intended to provide competition locally. It is my submission that this policy has been largely a failure, and in any case, needs to be radically reconceived in the light of stupid network, end-to-end arguments. Facilities-based competition is the idea that two, three and more service providers will reach a sufficient number of homes and businesses that the monopoly power of telephone and cable companies will be curtailed. The assumption in may cases was that the competitor was going to be a micro-telephone company, run on the same engineering principles as its larger rival. Enough of them have gone bankrupt that it is time to reconsider what competition would consist of in an Internet world, and where facilities-based competition would fit into this picture.

In a system such as the Internet, protocol interfaces are the new bottlenecks. Hence, when the Internet “interfaces” with a legacy network, the design philosophy of one network runs into the other, and the regulatory issue is: which one will prevail?

Unfortunately, regulators and policy makers are not sufficiently conscious of the opportunities afforded to legacy networks to frustrate the wealth-creation model afforded by the end-to-end principle

It is my purpose briefly to outline principles that would enable relevant action to be taken to protect and enhance the Internet from incumbent carriers’ strategies. To this end I propose a revised understanding of “facilities-based competition” that would adapt it to the circumstances created by the Internet’s protocol-based architecture. It will examine what is entailed by the concept of equal treatment of packets, and what relevance a slogan like “mind the protocol stack” would have for regulators acting on their traditional concerns for market power.

4. The Briefest Possible Comparison of the Internet to Telephony

When the telephone network computerized, analog equipment was replaced by computers without changing any of the fundamental design ideas and assumptions of the circuit switched network[11]. They are:

- the system holds open a circuit from end to end continuously throughout the call;
- the characteristics of the network are built around the usage patterns of humans and their hearing abilities;
- intelligence is scarce and is conserved by placing it inside the system, hence the "intelligent network";
- terminals accordingly are stupid, and are slaved to the nearest central office;

- the purpose of the system is to collect revenue; hence usage is tracked where revenue can be collected.

The point to be made about these design concepts is that no fundamental change accompanied digitization. The move from analog to digital changed the machinery but not the design of the public switched telephone system. Equally important, the assumptions of the telephone network remained firmly embedded in a pre-computer era.

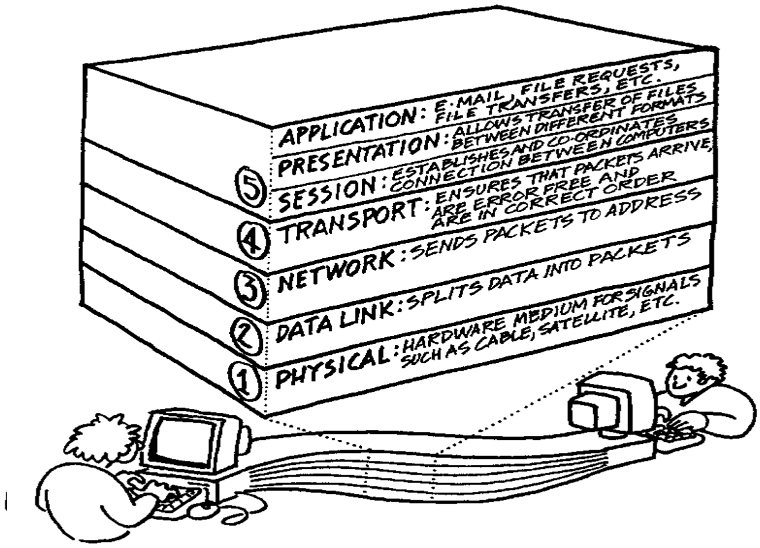
The telephone network is comprised of telephone switches linked together via telephone lines. The data, which travels on each of these digital telephone lines, represents the digital encoding of human voice at a rate of 64000 bits per second. At that speed the channel is called a DS0[12]. The function of a telephone switch is to set up and tear down DS0 channels and then to time the duration that the channel was kept open. From this information a phone bill is laboriously computed, which is ultimately passed on to the consumer.

Contrary to the PSTN, the Internet is not a "switched" network, but rather a packet-routed network. The Internet uses routers instead of switches to send packets to a particular destination. The difference between a switch and a router lies in the duration of the switching process. A router ultimately ends up making switching decisions for every packet, whereas a telephone switch makes a switching decision once for every call. The switch makes its switching decision based on an instruction coming from a control network (the Signalling System 7 or SS7 network) and the router makes its decision by looking at the headers inside the packet. A switch has to be warned that it is about to receive data from another switch whereas a router is constantly in a position where it is expecting data from other routers.

Feature	***Circuit-Switched***	***Packet-Routed***
Number of applications possible on the network	Only one: opening and closing the circuit	Potentially unlimited. The value is ultimately unknown unless each application is specified in the headers of the packet, which is never the case.
What can be measured (and thus billed for)	The duration that the circuit is kept open	Every piece of information identified in the header of the packet, the size of every packet and the percentage of link utilization
How it is controlled	By an external control network which instructs switches to set up or tears down the circuits	By link-by-link per packet forwarding control devices, i.e. routers
Knowledge by the system of itself (state)	High degree of self-knowledge	Low degree of self-knowledge

Traditionally, telecommunications was a single application, to apply a computer-era term anachronistically. Telecommunications policy makers are thus experts at regulating *single application networks*. Since the advent of packet-networks, the job of the regulator has become much more complicated as networks are no longer limited to one application.

The Internet's layered software architecture can be illustrated in the following way:



The Internet is characterized by:

1. A layered architecture that separates transport from applications by the TCP/IP layers;
2. An end-to-end architecture that puts the minimum of functions into the network and the maximum of functions into the terminal (computer host) - hence the title " Stupid Network";
3. A chaotic and adaptive routing pattern that sacrifices central control and predictability for maximum network efficiency, which results in ….
4. An absence of central ownership or common planning, as each network is privately owned, and may communicate or pass traffic at the owner's discretion; which results in the fact that…
5. the Internet only knows "autonomous system numbers", that is to say, networks. States and territories have no significance for how it works.

The telephone system is characterized by:

1. A vertical integration of transport functions with service, since the telephony concept predates the possibility of separating "applications" from transport, via computers;
2. The "service" is the product of the intelligence of the network, rather than the user's terminal, hence the "Intelligent Network";

3. The routing system is designed around the characteristics of the human voice to produce highly predictable results, based on the calling patterns of humans[1];
4. Central ownership of the network, common planning, and carefully controlled interconnection with other carriers, according to international standards, which leads to…
5. a system based on territories and governments, which work through state-based international agencies, such as the ITU to coordinate standards.

5. The Regulatory Tasks

For regulators, the chief issue lies in the interconnection of Internet-based architectures with telephony and the cable system. At those points of interconnection, one design must necessarily prevail over the other. Either we are to get the platform for innovation, the end-to-end principle, or we go backwards to the principle that "you get what we deliver".

Remarkably similar concepts underlie the legislation of other countries that regulate telecommunications carriers, since the problem to be addressed derives from the common characteristics of telephone technology.

In essence, the Internet has made the new situation much more complex, as protocols are stacked on top of one another. Opportunities for self-preference by the owner of the physical facilities have been multiplied.

We consider that three principles should guide the regulator in its discharge of its duties to introduce new and relevant forms of competition in previously monopolistic markets. They are:

[1] For example, if everybody simultaneously picked up their telephone, not everybody would get dial-tone.

1. facilities-less competition,
2. equal access for packets, and
3. minding the protocol stack.

Facilities-less competition

As a general rule, facilities-less competition means that whenever there is a choice to use less or fewer facilities to enable equally effective competition, the bias shall be towards any outcome which results in less or fewer facilities.

Equal Access for Packets

Equal access for packets is a governing engineering principle by which no one is allowed to exercise self-preference by any means. This might include any of the following: to configure, locate, load-balance, render redundant, cache, prioritize or manipulate by any apparatus or method, directly or indirectly, packets originating from a customer and intended for a competitor.

Minding the protocol stack

Minding the protocol stack is a governing regulatory principle that recognizes that opportunities for self-preference and undue or unfair discrimination can arise at higher levels in the protocol stack. Accordingly, regulators must take account of the arrangements of incumbents with their competitors at layers *above that of the physical* *connection of devices,* and explore the nature of the software by which the machines operate.

6. The Regulator's Problem

In North America, we have yet to see competition develop in residential telephony despite several decisions requiring its implementation, made more than four years ago. The current absence of residential competition is amplified by the fact that the focus of carriers, as regards residential service, has now completely shifted away from traditional circuit-switched telephony services towards high-speed Internet access.

Given this new focus, no new competitor will ever seek to build new facilities until such time as all opportunities to benefit from the existing infrastructure are no longer available. The outcome of existing telecommunications policy - which does not explicitly address the issue of high-speed Internet access in a uniform manner - prevents viable competition from emerging, because the relevant form of competition no longer consists of delivering circuit-switched telephony, but rather in allowing for all the possibilities of the Internet to reach customers.

At the moment the telcos seek to extend the utility of their copper pairs by means of DSL technology, but in the United States its roll-out is painfully slow and it is still inadequate to the task. On the issue of slow roll-out, an example will suffice. The Big Hook Conference was a selected gathering of Internet financiers, entrepreneurs, policy wonks, and geeks, held at Wood's Hole, Massachusetts, in September, 2001. When the fifty participants were asked: “How many of you do not have access to DSL in your neighbourhood?”, three-quarters answered ‘no’[13]. These are among some of the most tech-savvy people in the United States, living in places where DSL would be expected first. DSL is wholly inadequate for an always-on world of computers. DSL is a relatively narrow bandwidth and labour-intensive stopgap system which attempts to delay the need for deploying optical fiber to the neighbourhood and then to the home.[14]

Despite the inadequacies of DSL, we are not likely to get broadband networks any closer to the home until we overcome fundamental problems arising from the architecture of the Internet.

7. Investing in Infrastructure

New Internet applications can replicate the services provided with circuit-switched telephony and conventional television equipment at a fraction of their cost. For example, the Session Initiation Protocol (SIP) Internet protocol is presently being used to unify Internet Telephony, Internet Television and Electronic Commerce applications across a common set of TCP/IP interfaces which are far more powerful than what the Signaling System 7 will ever perform.

For as long as it may be possible to interconnect with incumbent carriers to provide high-speed Internet access services, the incentives to build new infrastructure will be ineffective. The costs of building a local access network are far too high in relation to any perceivable financeable return on investment, and this is a sufficient explanation. But the Internet exacerbates the investment problem by its separation of applications from transport.

Recall that in a telephone or cable system, the money is made on owning the services which are to be provided. Alternatives are not possible. Content and transport are integrated, and by making and then squeezing the choke-points, services can be sold at high prices indefinitely. The advantage to society of this arrangement was that networks were built. The disadvantages flowed from the monopolistic business arrangements that ensued. The Internet offers the chance to fix the problem of monopolies forever.

The Internet dissociates the applications from the transport by means of a code layer. The application layer, which is everything we ever obtain from computers, is intensely competitive. There is no surplus that an owner of businesses in the applications space can transfer to the bit-transport business. It is probably this realization that began the flight of capital from infrastructure investment in October of 2000. This is the Great Big Problem of the Internet, which capitalists and society must find a way to solve.

The solution that is beginning to emerge is to have local jurisdictions, such as municipalities, employ private companies to create optical fiber infrastructure. Alternatively, consortia of private users can build their own networks. This infrastructure would be available to all on an equal basis. Users would light their fiber according to the protocols that suited their needs. Common meet-me points would allow for interconnection with other users, with radio-based access technologies, and other carriers. The capital costs would be borne by local jurisdictions or by consortia of users. Existing large carriers could rent space and offer services, and save themselves the costs of establishing networks themselves. Capitalists would make money lending local jurisdictions the money to build the systems.

These municipally-owned systems take us back to the era in the 1920's when half of the United States was served by non-Bell telephone systems. They would differ fundamentally from the older architecture of circuit-switched telephony, however. They would be all-optical systems run on principles of open access to the infrastructure for all comers: carriers, private users, governments, anyone. There would be non-discrimination among users and uses. Regulators would oversee the terms of interconnection to these as well as to legacy systems.

Utopian? Such a system already exists and has been operating for several years in Stockholm, Sweden. Other experiments, such as Chicago's CivicNet, follow the same principles. In Canada and in several of the States, community-owned or privately-owned open-access networks have been created[15].

8. Conclusion

The principle social and political task for communications policy is to realize the benefits of the Internet by allowing it to prevail over legacy networks. The task involves regulation of the protocol interfaces between legacy and Internet-related systems, which alone should give regulators years of new tasks in exchange for the need to learn new things. The second and larger task is to gradually rebuild our networks in a manner consistent with the advantages offered by the end-to end principle, which is that users should drive the network, not network owners. It took a long time for North America to move from a railway economy to a highway economy, and to build the highways that were consistent with the possibilities of automobiles. It will take almost as long for us to move to networks consistent with the possibilities of the Internet. It may take a decision similar to that taken by President Eisenhower to build the interstate highway system, which was justified as a defence measure. However it happens, it is clear to me that we will not get there until we have a coherent picture of why the two legacy systems cannot deliver the social benefits of innovation made possible by the internet architecture. It will not happen as long as we pursue a complacent policy of "letting the market work" and other bromides with which we avoid the difficult work of actually thinking about the political economy of networks.

Endnotes

[1] http://www.crtc.gc.ca/archive/Decisions/1997/DT97-8.htm

[2] *A Brief History of the Future, The Origins of the Internet*, by John Naughton ISBN0 297 64330 4 , particularly Part II
[3] "End-To-End Arguments In System Design", J.H. Saltzer, D.P. Reed and D.D. Clark , M.I.T, Laboratory for Computer Science, Published in ACM Transactions in Computer Systems 2, 4, November, 1984, pages 277-288, and found at http://www.reed.com/Papers/endtoend.html
[4] The 7-layered OSI model was not really followed by the designers of the Internet. The session, presentation and application layers of the OSI model are compacted into one, leaving *applications* running over TCP (transport control protocol), the transport layer, which in turn runs over the network layer (Internet protocol), which runs over the data link layer, which runs through a physical transmission medium.
[5] J.H.Saltzer of MIT explained that this was a standard way to design computer functions at a conference at Stanford University on the end-to-end principle hosted by Professor Lawrence Lessig in October 2000. The point is explained in David Reed's "The End of the end-to-end argument" April 2000 at http://www.reed.com/dprframeweb/dprframe.asp?section=paper&fn=endofendtoend.html
"We agreed to architect the primary protocols of the Internet with only datagrams at the center. Vint Cerf and Jon Postel were persuaded to take a risk on a new style of network architecture, based on a radical decentralization of function."
[6] Lawrence Lessig of Stanford University documents the many attacks on the Internet as a platform for innovation in his *Code and other laws of Cyberspace*, ISBN 0-465-03913-8
[7] As David Reed said at the Big Hook conference, Wood's Hole, September 6, 2001 "We didn't design it in". See also *ISP Survival Guide*, by Geoff Huston, ISBN 0-471-31499-4, chapter 3, the Internet Protocol at page 41 and following, especially page 57,
[8] "The End of the end-to-end argument", by David Reed, April 2000 at http://www.reed.com/dprframeweb/dprframe.asp?section=paper&fn=endofendtoend.html
[9] See David Isenberg's "Rise of the Stupid Network" for an explanation of the difficulties in trying to innovate through the circuit-switched system <http://www.rageboy.com/stupidnet.html>
[10] As my colleague Rob Frieden des cribes in a paper for the 2002 PTC "Revenge of the Bellheads: How the Netheads Lost Control of the Internet".
[11] "Rise of the Stupid Network", <http://www.rageboy.com/stupidnet.html>
[12] The reason why 64000 values of either ones or zeros per second are necessary to transmit one second of human voice have their roots in the same science which could be used to explain you how music is stored on a digital Compact Disc. A digitization at a rate of 8000 representations per second is necessary to capture the highest pitch of the voice, which is 4000 Hz. The dynamics of human voice needed to be encoded using 8 bits of information in order to yield the same "toll" quality that people were accustomed to in the non-digital version of the PSTN. By multiplying those 8 bits by the rate of 8000 samples per second, we obtain the value of 64000 bits per second, which is the value of the DS0 unit used as the basis of capacity measurement on the PSTN. A link, which is capable of transferring 24 DS0 in parallel, is called a DS-1. The DS -1 is still used today as the principal unit to measure the interconnection capacity between two telephone switches.
[13] From personal observation at the time.
[14] Peter Cochrane, former CTO of British Telecom, and now head of Concept Labs, describes ADSL as "a product of the diseased minds of a telco. They have no understanding of how cross-talk affects it. Many people need to be employed to make it work. ADSL is a job creation program. Optical fiber eliminated 90% of the people in long lines, and fiber would do the same in local. Telephone companies should have given away cell phones, sold off the copper and rented out duct space." Notes from the Big Hook Conference, Wood's Hole, Massachusetts, September 5-7,2001.

[15] See, for instance, "The Coming Revolution in Dark Fiber Networks", by Andrew K. Bjerring and Bill St.Arnaud, Canarie Networks Inc. at http://www.canarie.ca/advnet/canet3/fibre.html