

Chapter 2. The Evolution of Internet Charging Arrangements

The major question this chapter explores is why the charging arrangements for the Internet differ so radically from those that apply in voice telephony. We ascribe several causes for this difference, labelled for convenience Technical, Regulatory, and Business. This chapter explores each of these in turn.

The chief fact that needs explaining is why connections among the networks that form the Internet is governed by private contract rather than by treaties or general obligations of law to afford interconnection under terms that are ultimately resolved by a regulator.

The difference between telephony and the Internet goes far beyond the contractual nature of business relations in one system and the regulated common carrier nature of the other. The abiding difficulty that telecommunications establishments have experienced in adapting to the Internet lies in the completeness of the difference between the two systems. It is not only the business relations among carriers that differ; it is the entire conception of the system.

At the heart of the Internet charging dispute is the perceived unfairness of payment systems that do not conform to the collegial sharing model of telephony, and the desire to make the Internet conform somehow to the more comfortable assumptions of the old order. The pressure from trans-Pacific ISPs and carriers to change the current system of charging arrangements, if it is to succeed, must be grounded in an appreciation of why the Internet has been designed to be so different from telephony. The current payments system was not foreseen by the founders of the Internet when they began working on computer-communications protocols in the 1970s, but it emerges naturally from a system that sought to place creativity and services in a different place than where telephony could ever provide it. The payments system, we shall argue, is an emergent property of the design of the Internet. Whether it can be changed while still preserving the creativity permitted by the current structure is a question worth asking.

I. Technical

Change at the rate of the computer industry

The starting point to understand the difference between telephony and the Internet is in the role assigned to carriers. Where do services come from? Who provides them? In the world of circuit-switched telephony, the answer was simple: it came from the carriers, who were also service providers. In the world for which telephony was designed, both switching and transmission were expensive, but switching, the intelligent function, was particularly so in a pre-computer era. Intelligence had to be conserved.

The way to do this was to concentrate functions in the network, rather than in the terminal. Exactly the opposite applies to the Internet.

In the Internet, intelligence is placed as much as possible in the user's computer, rather than in the network. This design philosophy was first presented by Salzer, Reed and Clark in paper in 1984,¹ and is called the *end-to-end* architecture. As it came to prevail, it resulted in a design where such decisions regarding the interaction between the host (computer terminal) and the network should, as far as possible, be delegated to the host, rather than attempt to supplant or duplicate the host functionality with network functionality. The overall design criterion was to make an absolutely minimal set of assumptions regarding the capability of each component network in the end-to-end path. From this design criterion emerged such features as "best efforts" routing and the omission of "state" – self-knowledge of the network. With the omission of state there is no concept of a "call". With no notion of a call, the routers merely forward a packet according to a global addressing system, and take no further interest in its success or failure to reach that address.²

By contrast, in the telephone system, intelligence of the system is placed in the network, rather than the terminal. The telephone handset is still "slaved" to the "master" in the central office. The use of the terms is not an exaggeration; it is a precise description of where the intelligent functions are located.

The concentration of intelligence in the network is the central design feature of telephony. It rests on assumptions about the cost of artificial intelligence that were made obsolete by the onrushing advance of computer technologies. However, the telephone system is and was protected by public policy from serious competition by a multitude of advantages that went with incumbency. A monopoly telephone system is a cash machine: it generates lawyers, economists, interest groups, lobbyists and politicians to defend it. It takes time, a competitive economic system, and the advantages of a technologically literate policy establishment – the least available part of the mix – to realize that there is no competitive advantage to be gained by preserving the telephone system from competition.

That competition takes two forms: the first being long distance and radio based competition (cellular telephones) and the second being more fundamental – a different system entirely – the Internet itself. It was not until the Internet that people began to see that competition to the telephone monopoly could take a much more radical form than replicating circuit switched telephony at lower prices, or greater convenience, which is what conventional telecommunications competition achieves. Competition could now take the form of another model entirely.

The model is of computers and software. Its features are:

¹ J.H. Salzer, D.P. Reed and D.D. Clark, "End-to-end arguments in system design", *ACM Transactions on Computer Systems*, Nov 1984, p. 277-288.

² Huston, Geoff *ISP Survival Guide*, John Wiley and Sons, Inc, 1999, ISBN 0-471-31499-4, pages 43-44
Savage, Frieden, Denton
International Charging Arrangements for Internet
Services - Module 3, Final Report

- Orders of magnitude declines in the prices of artificial intelligence;
- The ability of a multitude of software creators to devise products and services for common platforms.

These have the effect of imposing in turn:

- Relatively low costs on consumers to change, experiment or upgrade their software;
- Extremely rapid product obsolescence causing rapid life-cycles in firms, generating the requirement to innovate or die.

Once it was possible to link computers through the TCP/IP protocols, the issue began to be clear: which model would prevail, one drawn from the dynamic world of computers, or one based on the older model of telephony? The history of the last 20 years has shown that, so far, the computer model is prevailing.

We have argued that the Internet signals the arrival and ultimate dominance of the computer model, a dominance based in fundamental economic and technological characteristics.

One of the principal features of this new system is the lower cost of change associated with its software-driven nature. The cost of changing the Internet system is much lower than any comparable change in a telecommunications system. People buy or download software. If they like it, they keep it. If they do not, they discard it. Machinery – expensive physical objects – does not need to be installed and replaced. Hence change occurs on the Internet at the rate of adoption of software, not at the rate of change of expensive hardware. The cost of changing one's computer every few years is affordable and getting rapidly more so. The cost of ripping out all the circuit switches in the world and replacing them with routers, while still very expensive, is feasible. The cost of replacing the wires on which traffic moves can only be undertaken by the largest of enterprises over a period of decades.

The lowered cost of change allows for an increased rate of change. Consequently, when change is propagated on the Internet, such as a new service, change takes place like a viral infection. The cost of changing the physical plant of a network does not stand in the way of change. Product ideas like ICQ or Pointcast can arise, prosper, and disappear within three years.

The lowered costs of change and the rapidly increase rate of change point to the fundamental difference between the role of the network in an Internet world and what it was in the telephony era, which still continues. In a telephone network, the network defines what services are and what could ever be produced. Technological and service innovation was at the mercy of the telephone companies, with their need to extract all the revenue possible from their existing investments, their long depreciation schedules (approved by regulators wishing to keep telephone service affordable), and their monopoly. So the monopoly of the telephone system is in essence not over this or that service, but in defining what “services” are or could be.

The world of linked computers that we call the Internet radically devalues this business model. The network linking these computers plays fundamentally less important role than it was in a telephone era, where the network provided the services. In the Internet, the “services” are computer applications, and are bought and sold on compact disks.

A counterview

While this view may reflect what the Internet was up until 1998, or thereabouts, the Internet is not evolving in this direction any more. Integration of carriers and webhosters (the content that people want) means that a new dimension has been introduced. Innovation in software applications can continue rapidly without affecting this new feature, where the owners of content (or of access to the content, which is the same thing) are able to extract profit from those seeking access to it. Thus software innovation can continue unabated, without affecting the ability of network owners to extract economic rents from the ownership of content.

Some would argue that, by failing to mandate adequate open access to networks, regulators of telecommunications facilities (in which we include coaxial cable plant) are driving independent ISPs out of the access business and reinforcing the market power of those who control the facilities by which we access the products and services of the Internet.

It is not within the scope of this report to try to figure out the future of access policy. It is sufficient to say that nothing so far done in relation to the Internet in North America has slowed the rate of innovation in software, the dynamics of where creativity arises or how it is rewarded or the financing of new enterprises, nor has it lessened the contrast between the computer-Internet system and the older telephone model.

An example may suffice.

The World Wide Web

Very few people had heard of the Internet before the World Wide Web was introduced. Sometime after 1990, Tim Berners-Lee, then working for CERN in Europe, devised a software language for keeping text and illustrations together as they were sent over the then rather inconvenient academic packet system called Internet. With the Web came Web pages, “headers” – indexes of the content of Web pages, and then search engines, which were made possible by the indexing of Web page content. Before the Web, the Internet was an arcane instrument for scientists to communicate. After the Web was introduced, the Internet phenomenon took off. The domain name system (DNS) became necessary, for instance.

It is remarkable that the software that launched the Internet revolution was given away for free. No machine had to be physically modified to accept it. All one needed to do was download the software. A change of this nature would be impossible in a telephone system.

Traffic doubling every four months

Since the rate of change on the Internet is driven by advances in software products and services, which exploit the advances in computers, transmission systems must grow to match the vast new demands placed on it.

Machines know no natural limits on how much they communicate. Give them more bandwidth and people will find new ways of causing computers to use it, such as music and video. By contrast, a person has limited time in life to spend on the telephone, even if the price was free.

Internet traffic is doubling every four months, a rate faster than that at which computer power is doubling. Trying to provision enough bandwidth to keep pace with the Internet's growth is a major undertaking. John Sidgmore, the President of UUNet, a major Internet carrier, said his engineers told him, "If you're not scared, you just don't understand".³

The volume of Internet traffic surpassed voice telephone traffic in 1998. By the year 2008 voice telephone traffic will probably be less than 1% of Internet traffic.

To conclude:

- On the Internet, change is driven by the unregulated computer and computer software industry;
- Change propagates virally, as one user after another downloads the software;
- There is no natural limit to the extent to which computers will use bandwidth;
- Computer traffic has already surpassed voice telephone traffic. Eventually telephone traffic will essentially disappear in the noise.

II. Regulatory

While the computer model explains the rate of change on the Internet, regulatory decisions explain how telephone price regulation has never taken hold.

Four major decisions shaped the Internet and assisted Internet penetration in North America:

³ A link to this speech can be found at www.tmdenton.com

1. Declaring computer services “enhanced” and excluding them from “telecommunications”;
2. Having fixed price local calling areas;
3. Classifying Internet service providers as customers of the telephone company;
4. Keeping broadcasting regulation away from the Internet.

Enhanced/basic services distinction

In a sequence of rulings in the 1970s and 1980s in the US by the Federal Communications Commission (FCC) and in Canada by the Canadian Radio, Television and Telecommunications Commission (CRTC), “enhanced” services were excluded from the definition of “telecommunications” services. The need to do so grew out of the challenges posed by computers. When computation costs were high, computer service bureaux centralized computing in remote locations. As the telephone system became digital – computerized itself – means had to be found to limit telecommunications regulation to companies having “market power”. As the computer services industry was vigorously competitive, and showed none of the signs of economies of scale and scope characteristic of telecommunications networks, the enhanced-basic division prevented the spread of price and profit regulation to the entire networked world.

The enhanced/basic division was first imposed by regulators anxious not to spread telecommunications regulation to competitive areas of the economy, such as computer services. Later, “telecommunications” was defined in more recent legislation to be restricted to the carriage of traffic. Computer functions and services were kept unregulated.

Fixed price local calling areas

Internet penetration is largely a matter of price. The calling characteristics of Internet use are such that a line has to remain open between the terminal and the Internet service provider. North American telephone policy has always provided for the “free” local calling area. The ability to open a circuit through the legacy system to the Internet, and keep it open for a low monthly price, has been a major factor in stimulating Internet usage in North America. In other countries, some carriers are now offering Internet access rates on a similar basis.

Per-minute charges for Internet access are major impediment to making Internet service affordable.

ISPs as customers rather than carriers

Regulators in the US, and to a greater extent in Canada, have assisted in keeping the price of Internet access low and the market in access services competitive.

Internally to North American economies, many smaller ISPs depend on access to telecommunications facilities provided by larger carriers, many of whom are traditional telephone companies. Access by the customer to the Internet depends on the facilities of the telephone companies and the cable television companies, which can exercise market power to deny equal or equivalent access to customers on independent ISPs.

Regulators have tried to protect the competitiveness of the Internet access market, by treating ISPs as “customers” of the large carriers, rather than competitors .

The decision to treat ISPs as customers of the carriers has had important results, chief of which has been that ISPs continue not to be “carriers”, and to have rights of access to the carriers’ networks on a non-discriminatory basis. This means that carriers cannot give themselves privileges, such as giving their customers access to the Internet through their facilities on terms better than they allow their ISP customers. Though there are many difficulties to work out, the basic decision to treat ISPs as customers will continue to have positive effects in maintaining the distinction between underlying carriers and the service providers whose signals ride on top of those carriers (ISPs).

Broadcasting regulation kept off

Both the US and Canada have decided not to apply broadcasting regulation to the Internet. This has meant that the licensing system that lies at the core of broadcasting has not been applied to Internet service providers or to people offering services over the Internet.⁴

To summarize, regulatory decisions in North America were made early to keep the Internet and the services it offers unregulated, to keep carriers with market power regulated, and to keep the price of Internet access as affordable as possible.

III. Business

The history of the Internet shows that government fostering and sponsorship was essential at its beginning, but that governments, particularly the portions dealing with science policy that started it, divested themselves of responsibility. The Internet, once created, was launched into the existing telecommunications and computer legal and regulatory structure. Whether it would succeed or fail depended on that structure.

The Internet was the creation of the US government’s funding of communications engineers, and its development sponsored by the US and some other governments. The early research into packet-routed networks and the development of the protocols that

⁴ Review of the *Communications Decency Act* in the Supreme Court of the United States, and in Canada, CRTC Telecom Decision 99-xx

cause it to run were paid for by governmental research agencies, and conducted by engineering professors and their assistants in universities. In truth, the kinds of research expenditure and the agencies where this research was funded were very far from the main-line government departments and the attention of politicians. The Internet emerged from the backrooms, so to speak.

The most important business decision was to privatize the Internet. By the term “privatized” we mean that the functions of the packet-routed network usually operated by the scientific research funding arm of governments⁵ withdrew from service, and devolved these functions on one or more backbone providers who agreed to carry traffic. In those countries where governments incubated an academic packet-switched network, a decision was made in the mid 1990s to devolve the administration of these networks out of the government-university research world into the world of private enterprise.

The result of this privatization was that backbone Internet carriers entered a world where they carried traffic generated by others, essentially for free. The business was paid for by Internet consumers paying their Internet access provider a fixed monthly or annual fee. Smaller ISPs could load their traffic onto the large backbone providers through public NAPs.

The traffic regime was Sender-keep-all (SKA). Obviously this was disadvantageous to the receiver. Two results were produced by this absence of a settlement regime. The first was an increasing tendency of the larger ISPs to “peer” among themselves: to exchange traffic without charge as long as their traffic volumes were comparable. Peering involves the obligation to pass on the traffic of one’s peers and all the customers of those peers. It does not involve the obligation to pass on the traffic of third parties.

The second result was to cause the immediate emergence of a hierarchy of carriers. A stable SKA peering structure requires equality in the peering relationship. Those not able to bargain their way to the peerage on the basis of their traffic volumes or the value of their content were left with the obligation to pay their upstream supplier of bandwidth. The outcome could not have been different, and this was the origin of the ICAIS dispute.

With the explosion of Internet usage following the development of the Web, and the search engines that it made possible, it was no longer necessary or feasible to keep the management of the Internet in the hands of a self-selected group of Internet pioneers, however well-qualified and well-intentioned.

Institutionally, the move of responsibility for the Internet from academe to business changed the culture, the incentives, and the actors.

⁵ The National Science Foundation in the United States, the National Research Council in Canada, and so forth.

With privatization went commercialization. Incubation emphasized accruing positive networking externalities – the proliferation of the Internet. Government sponsorship helped players emphasize connectivity over profit maximization. As the Internet privatizes and commercializes, the profit maximization incentive grows and the tolerance for free riders drops. SKA served a promotional phase, and selective peering serves the now more compelling incentives to operate efficiently and profitably.

Another vital reason why the Internet has failed to generate a settlement scheme arises from how ISP is sold to clients. The monthly or annual fee paid by an Internet user to his ISP does not explicitly cover the cost of the provision of Internet connectivity from the client to all other Internet –connected customers. Once traffic has left the ISP's network, the funding burden passes on to the next ISP in the chain and so on until the end is reached. Inevitably, there is no requirement for an inter-provider financial settlement between the originating ISP and the terminating ISP.

In other words, throughout the Internet, where one network ends and another begins marks the point where one carrier assumes the costs and another one passes them on. In the case of Asia-Pacific carriers, that point has been established as some place or places on the US West Coast, greatly to their dissatisfaction.

In a telephony environment, the starting assumptions are different. A telephone company offers end-to-end service, and pays other telecom providers to complete the transaction. Costs are incurred among the players in the transmission linkage, and settlements occur to defray those costs.

In the Internet business environment, there seem to be two models: cash-free peering, and client-customer payments to upstream providers. Peering arrangements without financial settlements are only stable when the peers see some advantage to it. Absent some form of settlement-based funding of the intermediate level of ISP for its transmission costs, there will be – or already is – a strong tendency for a carriers of Internet traffic to become as large as possible, as fast as possible. Smaller ISPs will peer within their regions to like-sized ISPs, and become customers of the larger Tier-1 ISPs.

Whether this outcome is acceptable or not is the basis of the dispute about international charging arrangements for Internet services. Nevertheless, the outcomes are proceeding directly and naturally from observed and comprehensible causes.