This work is distributed as a Discussion Paper by the

STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH

SIEPR Discussion Paper No. 01-04

The Beginnings and Prospective Ending of "End-to-End": An Evolutionary Perspective On the Internet's Architecture

Paul A. David
All Souls College, Oxford & Stanford University

August 2001

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

_____

The Beginnings and Prospective Ending of "End-to-End":

An Evolutionary Perspective on the Internet's Architecture

By

Paul A. David
All Souls College, Oxford & Stanford University

First draft: 19 April 2001
Second draft: 27 July 2001
This version: 19 August 2001

This is a revised and slightly expanded version of the paper published under the title: "The Evolving Accidental Information Super-highway," in the *Oxford Review of Economic Policy*, vol. 17 (2), Fall, 2001: pp. 159-87.

Contact author: Prof. P. A. David, All Souls College, Oxford OX1 4AL, U.K.
Tel.: 44+(0)1865 279313; Fax: 44+(0)1865 279299
Email: paul.david@economics.ox.ac.uk or pad@leland.stanford.edu

ABSTRACT

The technology of "the Internet" is not static. Although its "end-to-end" architecture has made this "connection-less" communications system readily "extensible," and highly encouraging to innovation both in hardware and software applications, there are strong pressures for engineering changes. Some of these are wanted to support novel transport services (e.g. voice telephony, real-time video); others would address drawbacks that appeared with opening of the Internet to public and commercial traffic – e.g., the difficulties of blocking delivery of offensive content, suppressing malicious actions (e.g. "denial of service" attacks), pricing bandwidth usage to reduce congestion. The expected gains from making "improvements" in the core of the network should be weighed against the loss of the social and economic benefits that derive from the "end-to-end" architectural design. Even where technological "fixes" can be placed at the networks' edges, the option remains to search for alternative, institutional mechanisms of governing conduct in cyberspace.

Keywords: Internet economics, "end-to-end" design, network architecture, QOS, congestion, usage-pricing, interoperability standards, bundling of broadband services, competition regulation, governance

JEL Classification: L.15, L.96, O.33

The Beginnings and Prospective Ending of "End-to-End"

Paul A. David*

I. INTRODUCTION

Within the past decade everyday life in the economically advanced regions of the world has been touched and in some parts substantially transformed by the advent of the Internet. The sheer scale that the system has attained in so brief a time is breathtaking. It may be regarded as the largest artifact in the known universe: there are now over 100 million network hosts, and some 200 million PCs are connected on-line, and almost 30 million web-sites on the World Wide Web (Zakon, 2001). More than 100 million people in the United States were said to be using the Internet in 2000, to communicate with friends, family and colleagues, to access archived information in text and graphical images, and for electronic commerce. Western Europe as a whole is likely soon to be following Finland and Sweden's intensive deployment of computer-mediated telecommunications, and so will be closing the gap with the U.S. in terms of Internet penetration and the growth of business to business commerce conducted on "the Net."

The speed at which the numbers of host computers on the Internet have grown, like the pace at which global connectivity has been established, and the phenomenal proliferation of diverse innovations in applications software, are marvels that today win almost universal applause. With good reason, too, for they distinguish this communications infrastructure's performance from that of its historical predecessors – the telegraph and the public switched telephone networks. This is more than merely a further step in the expansion of communications capability. The unprecedented magnitude of the social and economic impacts that the Internet is exerting derives equally from the concurrent proliferation of powerful tools for exploiting the explosively expanding information resources to which effortless global access has been made possible.

It is the openness and transparency of this network – resulting from the distinctive "end-to-end" design of the architecture and transmission control mechanisms – that enables the Internet to tolerate extreme diversity and heterogeneity in the technical specifications of its constituent networks and platforms. Although the user perceives the Internet as though it were one single homogeneous network, in actuality it is a softly integrated heterogeneous network of networks. It attracts new network operators, service providers and users because they do not have to sacrifice proprietary solutions and idiosyncratic preferences in order to achieve connectivity. Connecting to the Internet does not require the using organisations to reshape the internal structure of their internal networks ("intra-nets') or install complicated gateways – except, of course, where security considerations dictate the use of information "firewalls" to block unwanted access by outsiders and the unauthorised export of data. Thus, the direct fixed costs for interconnection remain at a quite minimal level, as there is no need for extensive reconfiguration of the organisation's previously customised local-area

* *Senior Research Fellow, All Souls College, Oxford, and Professor of Economics, Stanford University.*

This paper would not have been attempted were it not without the background acquired through my exposure to the engineering, legal and social science expertise of the members of the Committee of the German-American Academic Council (DAAK) and the U.S. National Research Council- Computer Science and Telecommunication Board on "Global Networks and Local Values that committee. More specifically, I have drawn here upon my work with Raymund Werle in preparing a background document for the Committee (David and Werle, 2000). Private communications with Marjory Blumenthal, and the comments of Edward Steinmueller on the penultimate draft rescued me from errors on several points. In addition, I am grateful for the good advice supplied by Sheila Ryan Johansson, the editors of the Oxford Review of Economic Policy 17(2) special issue on 'The Economics of the Internet', Andrew Glyn and Andrew Graham, and an anonymous reviewer, all of whom sought to render my text more intelligible for non-specialist readers. None of those thanked here necessarily subscribe to the policy views expressed herein, and the responsibility for remaining deficiencies of analysis and exposition remains mine alone.

or wide-area network (LAN and WAN) to conform to new standards, nor to replace proprietary programs with Internet software.

An alternative, rather more lurid title for this essay might pose the question: "Will success destroy the Internet's end-to-end architecture?" That formulation has the virtue of suggesting the presence of a dialectical tension in the evolution of this communications infrastructure. "Success" in this context refers to the range of high-level features, or dimensions of system performance that are perceived by users of the Internet, -- i.e., the ease of effecting connection and of disconnecting from the infrastructure, its flexibility in accommodating users of heterogeneous hardware and software applications, the simiplicity of the interface for developers of novel applications that will run at the end-points of the network.. The source of that tension, considered in its most general aspect is simply that many of the performance capabilities that users of this telecommunications infrastructure value as uniquely beneficial rest upon certain technical specifications of the system that were not introduced with a view to meeting the system performance properties and service characteristics of users of today's Internet.

Those same technical specifications also give rise to other, more problematic "service characteristics" that have become manifest in the course of the surprising, spectacularly successful global deployment of this "unmanaged" network-of-networks. Consequently, economic and political forces currently are channeling engineering efforts towards providing remedies for the perceived drawbacks, as well as enhancements to satisfy the technical requirements of contemplated new Internet applications. These "adaptive modifications" cannot be construed to be part of some automatic, global process of technological optimisation. Quite the contrary, for, what is underway essentially is a decentralized, goal-seeking evolutionary dynamic driven by the interests of particular groups of Internet stakeholders. This process continues to draws support from the fusion of liberal individualism and technocracy in the philosophical-political ethos that has become quite pervasive among the community of Internet engineering specialists, and which is predisposed to reject social and legal modes of regulation in favour of finding purely technological mechanisms to address deficiencies in system performance. Add to this the real obstacles to negotiating any other arrangements for the governance of an unmanaged system that extends across numerous national jurisdictions, and it will be evident that establishing meaningful policy guidelines for the evolving Internet will be far from easy. Yet, in the absence of some concerted initiatives towards that end, it is not only conceivable but increasingly probable that the piecemeal introduction of new technical mechanisms in the core of the network will soon begin the destruction of those performance capabilities hitherto have constituted some of the Internet's most beneficial public goods properties.

II. THE STRUCTURE OF THE ARGUMENT AND ITS POLICY IMPLICATIONS

The "technical specifications" whose prospective modification is the subject of particular concern are those found in the layer of technology that controls and manages flows of data in the Internet, routing information between communicating applications that are located at the end-points of the network. The design and placement of these "bearer/network protocols" evolved from the solutions to the antecedent engineering challenges that, during the early 1960s, attracted the attention of several groups of digital telecommunications systems researchers. A variety of technologies were created in that era for inter-connecting computers over packet-switched data networks, but those which came to distinguish the Internet reflected the institutional settings, and the organisational needs of the design groups that pioneered their implementation in the ARPANET. Similar institutional conditions continued to influence the direction in which "inter-networking" tools used in the ARPANET were elaborated and eventually came to be deployed on the NSFNET. The influence exerted by the essentially stable "social parameters" of the academic research environments within which the distinctive technical specifications of these precursor networks were developed can be displayed clearly by tracing the historical genesis of the Internet's principal constitutent elements: the packet-switched data network's end-to-end architecture, the TCP/IP protocol stack (specifying the

transmission control mechanism and domain name system of the Internet addresses), and the features of the widely used network services, including email, file transfer protocols (FTP), and the WWW browser.

In the early 1990s, the operational backbone of the network constructed for the National Science Foundation (NSF) was rather unexpectedly opened to commercial traffic, and, shortly thereafter, an enlarged NSF backbone was transferred from government ownership to private hands. Thus, after two decades of development under public auspices and sheltered application in scientific research settings, where individual and organisational behaviors were regulated tightly by social norms and institutional rules, the Internet suddenly was released into a very different, and largely unprepared environment -- the world of telecoms users at large. An apt conceptualisation of the subsequent phase of the Internet's evolution therefore is that of the career-course of a wonderful, yet rather anomalous "found object." Far from being the instrument purposefully designed to effect a dramatic augmentation of the global communications infrastructure, the Internet is one of those fortuitous legacies of public programs of exploratory R&D, which different elements of modern society subsequently have become obsessed with trying to use in a variety of ways for which it was never intended and remains awkwardly suited. (See CSTB, 1999a, esp. Ch.7, on the role of U.S. federal government agencies.) In this respect the description of the Internet by *The Economist* magazine (July 1995) as "the accidental [information] superhighway," is fully justified by circumstances of this technology's early historical development, and also by the circumstances surrounding the manner in which it was made available to the world.

The decision to open the NSFNET backbone network to commercial traffic and use by the general public soon exposed some features of the Internet that appeared decidedly "dysfunctional" in this new environment. Prominent among these drawbacks were the difficulties of preventing the delivery of unwanted or offensive content; of ascertaining the identities and physical locations of either the recipients or the initiators of messages – including those responsible for malicious actions such as "denial of service" attacks; and of pricing usage to reduce congestion. Further, and related to the foregoing, the platform offered by the Internet was found to be ill-adapted for use by commercial enterprises that expected to use the familiar "fee-for-service" business model.

Contemporary efforts to address a number of these, as well as other problematic aspects of the Internet are being channeled towards technical engineering modifications which would facilitate monitoring and filtering of message content by public authorities, or by ISPs at access points created within the network. The impetus to seek such "solutions" more-or-less to the exclusion of other, non-technological approaches to amelioration, derives in some part from doubts about the efficacy of purely national regulatory responses. Such hesitancy is quite understandable, considering the obstacles that the global reach of the "network of networks" now posed to establishing effective transnational governance of cyberspace.

But, the implicit policy preference for technological solutions also has been reinforced by more explicit, and in some quarters vociferous hostility to the exploration of any legislative or legal remedies. The historical circumstances surrounding the privatisation in the early 1990s of NSFNET, the Internet's immediate precursor, may be seen to have contributed to the formation of this current climate of opinion affecting what might be called "the political economy of the Net." The regulatory *status ab initio* that the NSF's decisions created was one in which the only governance institutions having responsibility for Internet matters *per se* were exclusively concerned with the engineering aspects of the network. Although that perhaps is not the most economically or socially significant among the path-dependent *sequelae* of the terms under which the Internet's career as a global information infrastructure was launched, ample scope thereby was left for resistance to virtually all save purely technological approaches to regulation. Proposed administrative or legal interventions by public authorities continue to meet forceful resistance. But this derives less from generic *laissez faire* sentiment within the business community than it does from subscribers to the political-philosophical argument that such actions would attack the very "essence of the Internet" – construed as a global interaction space free from governmentally imposed structures of *social* regulation, indeed, in more

extreme formulations, from the interposition of all authority that would circumscribe the users' freedom of action in cyberspace.

Developmental work on technological modifications that could alter the architecture of the Internet is gaining further impetus currently from the prospective availability of greatly expanded (broadband) capacity for offering new commercial services, particularly, voice-telephony and real-time video. In addition to consuming comparatively larger quantities of bandwidth, these services are intolerant of perceptibly long delays in transmission ("latency" in the language of specialists). Such delays, however, are a characteristic of the Internet's "best effort" approach to the step-wise forwarding of data packets between the sender and the host computer on which they are reassembled for collection by the designated receiver. Hence, would-be vendors of those and other, complementary services see profitable opportunities in the construction of proprietary islands and archipelagos on the network where "higher quality" telecommunications standards could be provided for their customers. The deployment of some technologies that would provide differentiated quality of service (QOS) constitutes a plausible evolutionary path along which the ending of "end-to-end" architecture may be driven by private business initiatives, rather than by technological modifications introduced to enable governmental control functions.

It is in just this connection that another among the numerous unexpected consequences of the Internet's deployment had assumed considerable potential importance for policy-setting: the opening of interconnection to public telecommunications domains (cellular radio, satellite and cable) into which the providers of commercial services can migrate. By doing so, some providers of Internet services (ISPs) may be able to escape from existing legal and administrative restraints that, for historical reasons, had been imposed upon businesses based upon other communication modalities. The implications of this facilitation of "regulatory bypass" are illustrated by the current asymmetric regulatory treatment of the telephone industry and the cable broadcast industry in the US. Network operators in the long-regulated telephone who offer broadband access to the Internet have been required, largely for reasons of competition policy goals, to provide their customers with open and non-discriminatory access to other broadband ISPs. Cable companies, although performing the same functions, find themselves under no corresponding regulatory constraints. As a result, informed observers recently have expressed alarm that the existing regime of regulation (and non-regulation), by permitting the cable companies to bundle broadband access with selected application services offerings, is creating particularly powerful private economic incentives for what might be called the "business balkanising" of the Internet (see, e.g., Lemley and Lessig, 2000).

There is no *a priori* reason to suppose that the most efficient path to "enhanced" Internet performance is one that relies solely upon "fixes" and enhancements that can be implemented in computer hardware and software. Introducing engineering solutions into the core of the network would entail a sacrifice of the future benefits provided by the "end-to-end" architecture, and thus may not be justified by the social value of the "performance improvements" thereby achieved. Where it is not feasible to place technological solutions at the edges of the network, there is at very least a potentially strong case for directing greater resources towards developing effective political and legal institutions to regulate the behaviours of Internet users. Rather than being viewed as substitutes, technological and social mechanisms of governing cyberspace may offer complementary solutions that should be explored in a coordinated, and necessarily multi-disciplinary fashion. When political and legal devices appear likely to prove unworkable, or excessively costly to maintain, and where the technological implementation of new functions would require sacrificing the Internet's "end-to-end" engineering, the rational policy course is then to acknowledge the existence of a conflict between alternative *desiderata* in the Internet's capabilities. The social value of the functionality that would be added to the performance of the global information infrastructure in that case should be weighed carefully against the benefits that would be lost with the ending of "end-to-end".

In the remainder of this essay some documentation is supplied for the main empirical propositions on which foregoing argument rests, and its principal policy contentions are unpacked for closer examination. Section III briefly recalls some history of the Internet, focusing upon the non-

commercial, homogeneous social contexts of invention and implementation within which the key enabling technologies were created and incorporated into the designs of the major inter-networks supporting communications among scientific research communities. The view of the Internet as evolving subsequently from its initial status as a marvellous but in some ways very awkward "found object", receives further elaboration from an examination of the circumstances in which the NSFNET backbone network was opened to commercial traffic at the end of the 1980s, and soon thereafter, was transferred to private ownership. This highlights some of the enduring and quite predictable consequences that followed from the unregulated condition in which the backbone of the Internet was privatised: the ensuing oligopolistic structure of the backbone industry, and the conduct of its dominant firms affecting the provision of capacity at key network access points (NAPs).

Section IV reviews the several emergent features of the Internet's performance as a general-purpose communications infrastructure that presently are perceived to be socially and economically problematic, indeed, in some instances seriously "dysfunctional." The discussion indicates how such problems have emerged to become focal points for proposed remedial technological modifications that would be implemented in the core of the network. An effort then is made, in Section V, to provide an expanded framework in which to consider the policy issues raised, on one side, by calls for improved mechanisms of governing behaviour on the Internet, and, on the other side, by the quest for new private profit opportunities. The benefits of the solutions proffered in response to these dual challenges must be assessed against the costs of the likely irreversible damage which could be done to the Internet's "end-to-end" architecture by a *laissez faire* stance that liberally accepted all technological solutions – so long as they could be shown to "work." Two generic technical questions are identified as being important in this connection. The first is whether the perceived need for a technical "fix" reflects a condition that otherwise would persist on the Internet, or whether the problem is likely to be transient in nature. The second is whether the proposed technological solution itself is likely to remain efficacious over the longer run, instead of providing only a temporary palliative.

The paper concludes in Section VI, by setting forth a few broad considerations that ought to guide policy approaches to controlling the character of technical enhancements and regulating human behaviours on the Internet, through the design of institutional mechanisms of governance as well as technological engineering solutions. This brief discussion underscores the need to take account of the present dynamism of the Internet's enabling technologies, and also of the limitations imposed by the distinctive architecture that was inherited from its historical precursors. It suggests, further, that for the field of "Internet economics" to mature into an area of disciplinary specialisation that has more immediate policy relevance, economists will need to develop a greater appreciation of both of those realities, as well as of the historically contingent processes of technological and institutional co-evolution.

III. THE DEVELOPMENTAL CONTEXT OF THE ENABLING TECHNOLOGIES

A significant element of historical irony is present in the circumstances that gave rise to the key component technologies of "the network of networks." We perceive the Internet today as affording global connectivity and access to a wealth of diverse information resources for a variegated multitude of organisations and individuals, an assembly of actors motivated by a wide variety of interests that are aligned in some respects, while conflicting in others, and guided by value systems that coincide on some issues but diverge strongly on others. It appears only too evident that human interactions in such a sphere would be bound to create the myriad problems of governance that arise in any reasonably complex social setting. Further complicating matters, the parties brought into contact with one another "on the Net" are not embedded within a unified social order, but, instead, are drawn from a culturally and politically heterogeneous collection of societies. Yet, the architecture of the network and the specifications of its components and service functions were not designed with anything resembling the present area of applications in mind. As may be seen from even the briefest sketch of the

developmental context of the Internet's enabling technologies – from the pre-ARPANET origins of packet-switching in the mid-1960's, the design of the TCP and IP protocols for host-to-host communications, and through the succession of widely used "network service" technologies from E-mail and FTP, to the WWW browser that CERN released for public use in 1991 – the most powerful persistent considerations shaping the developmental process were the communications needs of an entirely different, and far more homogenous institutional and social environment. This was the distinctive epistemic culture of the successive publicly funded scientific research groups that were themselves responsible for the invention and implementation of the Internet's precursor networks. In that sense, it might be said that the creation of the Internet was a reflexive technological achievement.

Although widely distributed geographically, and situated in a variety of academic and quasi-academic institutions, the technologists who took the lead in "casting the Net" and "weaving the Web" were very much alike in regard to the general ethos of cooperation and the social norms characteristic of their respective scientific and engineering work cultures. As will be seen, the membership of those groups consisted largely of scientists and engineers engaged in cooperative research undertakings, for which they had been recruited by reference to criteria of technical competence (and national security considerations, in some early instances). Beyond this considerable measure of social and professional homogeneity, the access of the members of these work-groups to the novel communication network they were fashioning initially was quite constrained; and the behaviour of individual users was subject to supervision and non-technological regulation by the respective university- and public institute-based organisations within which they were employed.

Quite understandably, then, the researchers who invented and improved what were to become the distinctive technical features of the Internet made few if any engineering provisions to cope with issues of content, privacy, security, identity, and so forth. Nor was attention given to devising means of protecting the functionality of the network from being degraded by attacks that could originate in the behaviours of its users, rather than from external agencies. This was so primarily because those issues were not problematic for the original designers, their institutions, or their sponsors.

Numerous historical narratives have detailed the contributions made by particular individuals and agencies to the technical development of the Internet. However, in order to see how the design choices of these pioneers has been shaped by the institutions of which they were part it is helpful to identify four distinct phases in the evolution of the Internet.[†]

(i) Data Communications and the Dawn of Packet Switching

The opening phase took place during the 1960's, when individual researchers based at different universities and research institutions in the U.S. and Britain were concurrently developing the underlying means to enable data communications between computers, including queuing theory, packet-switching and routing. A guiding conceptual framework, indeed, the first recorded description of the social interactions that could be enabled by computer networking, was set out in a series of memos written in 1962 by J. C. R. Licklider, a psychologist at MIT, who soon thereafter became the first director the Information Processing Techniques Office (IPTO) at the U.S. Defense Department's Advanced Research Projects Agency (ARPA). Licklider and Clark's (1962) "Galactic Network" concept envisaged a globally interconnected set of computers through which humans could quickly access data and programmes from any site. This phase was brought to a close in December 1968, when ARPA issued a contract to the firms of Bolt, Beranek and Newman (BBN) of Cambridge, Mass., to build a packet-switched network along the lines of the initial design presented in 1967 by Lawrence Roberts, an IPTO program manager recruited from MIT.

[†] The material on which the following paragraphs draw includes recent retrospective accounts provided by leading participants (Cerf 1997; CSTB, 1999a: Ch.7; Leiner *et al.*, 2000, Berners-Lee, 1999), as well as secondary sources, including Abbate, 1999; Hafner and Lyon, 1996, Hauben and Hauben (1997); Rogers, 1998; Salus, 1995; Zakon, 2001).

Most of the early research on networking that was encouraged by Licklider, and the subsequent directors of IPTO who he influenced, had focused on packet-switching. A paper written by Leonard Kleinrock (1961) , also at MIT, first proposed this technique, and the appearance of Kleinrock's (1964) book on the subject attracted still greater attention, because it offered an efficient means of handling the "bursty" transmissions that computers would generate. Unlike voice conversations which can be characterised statistically, computers communications are sufficiently unpredictable that were they to be carried through the connections that served voice telephony, it would become necessary to provide wide margins of extra (and only irregularly used) capacity in the lines. Given the high fixed costs of computers at the time, networking was an economically attractive prospect – as was time-sharing of mainframe capaciity – so long as it permitted attaining high utilization rate on this equipment without requiring the addition of expensive margins of excess transmission capacity.

Packet-switching, developed largely independently by IPTO, by Donald Davies and Roger Scantlebury at the National Physical Laboratory (NPL) in Middlesex, England and by Paul Baran (1964) and his co-workers at the RAND Corporation, was the means of achieving this. The breakthrough entailed a means of using the existing capacity of the PSTN infrastructure more intensively, by dividing streams of binary data (messages) into small units contained in envelop-like "datagrams," each of which carried the address of the final destination (host computer) to which it would be routed through the network independently of the others, until they arrived for re-assembly at their common destination. The researchers at NPL coined the term "packet" in referring to these small units of addressed data. Specialised computers -- Interface Message Processors, or IMPs, as these were dubbed at the time – handled the tasks of receiving and forwarding packets, both keeping track of the traffic and varying the routes taken by the packets to avoid congested nodes and failed links, functions performed today by "routers" on the Internet. When this design was first introduced, its critical feature was seen in freeing the scarce data processing resources of the mainframe "host" computers from network operation tasks. But the resulting separation of the (IMP-) computer mediator communications network from the data processing "intelligence" represented by the hosts at its edges was a radical, defining step in the genesis of the end-to-end architecture of the modeern Internet.

(ii) The Age of the ARPANET: New Network Services, the Genesis of TCP/IP

The second phase of development began in 1969 with the installation of the four initial nodes of the ARPANET. That this network was to serve as a research tool for host-to-host communications in support of the work of the research organisation that built it is apparent from nature of the projects at those nodes: (1) Leonard Kleinrock's research group at UCLA, (2) the Stanford Research Institute project, in Menlo Park CA, where Doug Englebart's On Line System (NLS) was used as the host, (3) Ivan Sutherland's pioneering computer graphics group at the University of Utah, and (4) a U.C. Santa Barbara project that was developing an interactive system for mathematical education (see CBST, 1999: pp. 172-173, 228-232). At first, the ARPANET was used primarily as a facility for experimentation with packet-switching, rather than a communications service for the researchers situated at its nodes. This necessarily was the case at the outset because the first protocols for host-to-host communications – the Network Control Protocol (NCP) – were not completed by the Network Working Group (NWG) led by Steven Crocker until the end of 1970.

Once the ARPANET sites had completed implementation of NCP during 1971-72, however, network users began developing service-applications. The Telnet protocol was introduced to allow a user on one machine to log onto another, at a remote site, thereby sharing ARPA's costly computing equipment; the File Transfer Protocol (FTP) also came early, in 1971, to allow a user on one system to connect to another in order to send or retrieve particular files. But the service that turned out to be the most the popular, the "killer app" for the ARPANET community, was the e-mail system (READMAIL) introduced in 1972 by Ray Tomlinson, who modified an electronic communications facility which he had previously developed (for users of BBN's time-sharing system, Tennex) so that it would run on the new network.

Although it breaks the chronological ordering of this narrative, it is nonetheless useful at this point to jump ahead to the circumstances that gave rise to the creation of the World Wide Web (WWW) browser, the new inter-networking service technology developed during at the end of the 1980's by Tim Berners-Lee, and which became the "killer application" whose popularity was a major factor driving the growth of the Internet from the early 1990's onward**.** The purpose in doing so is not simply to notice the general parallel with the previous case of e-mail on the ARPANET, but to underscore the aspect of similarity concerning the non-commercial, research supporting motivation in the invention of the WWW by Tim Berners-Lee and Robert Cailliau at CERN, the high-energy particle physics facility in Geneva, Switerland (see Berners-Lee 1999, CSTB 1999a, Hameri and Norberg 1998, Naughton 1998). The idea of the Web itself is simple enough: provide a uniform format for archiving documents stored on server computers, and assign a unique name to each document so that it can be located and retrieved by a browser programme. Because the unique names (termed universal resource locators, or URLs) would include the domain name system (DNS) name of the host on which the documents were stored, they were long and the URLs could more conveniently be represented as briefer hypertext links in other documents – thereby making use of the hypertext system that had been invented by Douglas Englebart back in 1967 (in a project that, incidentally, was drawing support from ARPA).

There was an essentially parochial impetus behind Berners-Lee and Cailliau's development of a document format for this purpose -- the variant of the Standard Generalized Markeup Language in use in the publishing industry since the 1950s, which they called the Hypertext Transfer Protocol (HTTP) and released as a new Internet protocol in July, 1992. CERN mobilises the work of thousands of physicists from many countries, who are organised in teams that cooperate to build and operate large experimental facilities; coordination of this work demanded reliable and rapid access to the documentation of all the parts of the complex apparatus that specialised groups were working on. This was the internal need that persuaded CERN to underwrite Berners-Lee to work on his proposal in the late 1980s. But, like e-mail, some two decades earlier, the resulting browser technology turned out to have a considerably wider field of application.

Coming back now to the 1970's, one must emphasize that of ultimately far greater significance than the introduction of novel "network services, " this period of intentensive research on the ARPANET saw the emergence of the technical design principles and specifications for a "connection-less" telecommunications infrastructure. These remain the fundamental features that distinguish the Internet from other, "connection-oriented" systems such as the public switched telephone networks (PSTNs). The key underlying technical idea was "open architecture networking," which Robert Kahn introduced in 1972, shortly after coming to ARPA from BBN. This approach called for the designing of a encompassing, meta-level "inter-networking" architecture which, instead of tightly dictating the specifications of individual network technologies and the interfaces between them, would leave such features free to be set so as to meet the particular performance requirements of their respective providers, while enabling them to function in conjunction with other, similarly unconstrained networks (see Leiner et al., 2001).

Kahn started with the task of linking a packet radio network with the wired ARPANET by means of a transport protocol that would be able to operate no matter how unreliable were the underlying links – due to conditions such as radio interference, due to natural terrain or to jamming, and which therefore could not be embedded in the infrastructure of the transport layer itself (CSTB, 1997: p. 20; Leiner, *et al.*, 2001). ARPA was engaged at this time in research on other applications of packet switched communications, including both terrestrial packet radio, and packet satellite transmission. Its' group working in Hawaii (Aloha project) was particularly interested in being able to interconnect computers by radio, rather than having to set up costly hardware on the islands. In 1973, working with Vinton Cerf (then at Stanford University), Kahn moved on from this particular problem to address the generic problem. The solution that emerged was the Transmission Control Protocol (TPC), which assigned to the sending nodes the responsibility for regulating the flow of packets in response to indications of network congestion based upon the cumulative acknowledgments from (adjacent) receiving nodes. It also specified an addressing mechanism that could accommodate as many as 4 billion hosts. Unlike

the ARPANET's NCP which enabled communication between hosts on a single network, the TCP specifications (separated into TCP and IP protocols in 1978) were designed to interconnect multiple networks, and so allow communication between computers on a variety of different networks (CSTB, 1999aa: p. 174.)

From the narrow technological angle the ARPANET (and its successors) may be regarded simply as one among a number of technical solutions to the problem of interconnecting host computers of different types, and, more importantly, also networks based on different switching and transmission technologies. In the late 1960s and early 1970s alternative technical solutions were being explored or were available already to interconnect the technically diverse assortment of data networks that had sprung up in the USA, including the Palo Alto Research Center (PARC) Universal Packet from Xerox, the Unix-to-Unix copy protocol (UUCP) originally developed in the Bell Labs of AT&T, which was licensed out practically for free, and the standards proposed by the ambitious OSI-project that was being promoted in Europe. (Hafner and Lyon 1996; Salus 1995; David and Werle 2001, for further discussion.) Consequently, at the beginning of the 1980s it was hardly a foregone conclusion that the open-architecture design and TCP/IP would emerge as the dominant technologies for inter-networking computers. Indeed, it was not until January 1 of 1983 that the TCP/IP protocol stack -- whose details had been published almost a decade earlier by Cerf and Kahn (1974) – actually replaced the original (NCP) control protocols on the ARPANET. By the early 1980s the ARPANET was supporting a number of operational defense organizations as well as R&D organisations, and the mandated transition to TCP/IP provided an opportunity for the DoD to split off a MILNET that would support operational requirements, leaving ARPANET supporting only the needs of the R&D activities.

(iii) NSF and the Rise of Academic Networking in the 1980's

During the 1980s new, higher capacity cooperative networks were more widely deployed to support university based research, a development that began in 1981 with BITNET, and CSNET; the latter (Computer Science Network) drew funding support from the NSF for the purpose of connecting those computer science departments that had no access to the ARPANET and consequently lacked not only sophisticated facilities for collaborative experimentation, but email communication. At its peak, CSNET had approximately 200 participating sites and international connections to fifteen countries. In 1987 it was merged with BITNET to form the voluntary, self supporting non-profit Corporation for Research and Educational Networking (CREN); with the development of the NSFNET the needs that had given rise to CSNET no longer existed and by 1991 the service had been discontinued Eventually, still more geographically extensive regional computer networks were linked in the U.S. and served a widening circle of scientific disciplines, requiring increasing technical and organisational attention to the provision of capacity and interoperability. Overseas, the European Academic Research Network (EARN) was established in 1983, and in the following year Britain launched the Joint Academic Network (JANET) program – the first network whose announced purpose was to provided interconnecting computing facilities for the entirety of the country's university research community, without regard to discipline. The National Science Foundation followed suit with the NSFNET program in 1985.

The NSF's immediate purpose in launching the NSFNET was to connect its five super-computing centers (and the National Center for Atmospheric Research) with university computers via what was at the time a "high-speed" (56kbps !) network, and this had been accomplished by 1986. But once it had been implemented as a general-purpose network, NSFNET began to serve as the "backbone" or upper tier of a hierarchical network of networks. Its ability to interconnect readily with sub-networks in this manner resulted from a critical event in 1985: following a proposal 1985 made by Dennis Jennings, who recently had come from Trinity College, Dublin, to lead this initiative at NSF, a crucial decision was taken to base the NSFNET system on the TCP/IP data communications protocol.

This was a controversial and hotly contested issue at the time (see Rogers 1998), since the main users of the super-computers were physicists and chemists who were accustomed to working with

DECNET, and protocols derived from other proprietary mainframe networks, and so were dubious about the wisdom of Jennings' proposal especially as there was no actual experience with super-computers using TCP/IP. On the opposite side were arrayed other, regional and local research and education networks whose participants were drawn from computer science groups that had gained experience with TCP/IP on ARPANET. Still others were pressing NSF for expanded access to data communications of the sort that TCP/IP would permit – by setting up connections to the NSFNET for local campus LANs, private data networks maintained by corporations engaged in research on computer science and networking, and regional WANs linking departments and laboratories at different universities. Networks of that kind evolved in the second half of the 1980s. Many of them were co-sponsored by private business organizations and this contributed to increasing their technical heterogeneity, as researchers in those settings were using LAN-technologies (such as Ethernet), X.25, SNA, DECNET and other proprietary solutions. Thus, BARRNET, the Bay Area Regional Research Network, included local research-oriented corporate members from IBM, Hewlett-Packard, Xerox PARC, as well as Stanford, the University of California at Berkeley, and the Lawrence Livermore Lab, with the non-university members being restricted from using these networks for commercial purposes other than research (see Headley 1995).

All these local and regional networks were required to adopt TCP/IP in order to become connected to the NSFNET backbone, and this had a significant impact in diffusing acceptance of the TCP/IP protocol stack more widely as the de facto national standard for inter-networking. Of course, at least three positive features of TCP/IP that entered into the NSF decision also were more generally conducive to its rapid emergence as a *de facto* standard. These were: (a) its use on the ARPANET, where it completely replaced NCP at the beginning of 1983, had thoroughly demonstrated the ease with which it could be used to interconnect heterogeneous networks, (b) the fact that, being a non-proprietary standard, it was made available through electronic distribution free of charge, and (c) TCP/IP already was part of the tightly integrated networking capabilities that were included in the U.C. Berkeley-produced version of the UNIX operating system (BDS). The latter, which ran on both DEC's VAX line of inexpensive (mini-) computers popular with scientists, and the new SUN personal workstations (from the fledgling SUN Microsystems, Inc.), at that time was gaining exposure and popularity among computer scientists.

Thus, at the end of the 1980s the TCP/IP supporters were growing in number and gaining acceptance from their previously sceptical peers in the private computer and networking industries, as well as from IT specialists in the mission- oriented public research labs of the Department of Energy and NASA. By the early 1990's, the momentum acquired by the campaign for TCP/IP within the technically sophisticated networking community had imparted to the nascent Internet some of its appeal as a "grass roots" movement that had been able to triumph in the face of resistance, and even initial hostility on the part of those steeped in the engineering orthodoxy of the PTT's telecommunications establishment. There is a connection worth noticing here, between the pioneers' shared experience of successful technological insurgency, and the reinforcement of certain technocratic traditions and attitudes that continue to colour policy discussions concerning the governance of the "unmanaged" global (inter-) network to which the success of TCP/IP gave rise.

What stands out in this, indeed throughout the first three phases of the development of inter-networking that have been reviewed, is the interplay between the institutional environment and the behaviour of the new technology's user-innovators. Early on, computer scientists formed particular communities, initially around specific time-shared computers, and later around programming languages, operating systems and computer networks (Norberg and O'Neill 1996). However, by the late 1970s and early 1980s there had emerged a broader sense of participating in a community that was pioneering revolutionary changes in information processing freed from constraining network architectures. This manifested itself in a more unified spirit of collaboration, informality, and even a sense of social responsibility reflected by norms and rules concerning the use of networks, that was especially striking to observers from other cultures (Leib and Werle 1998).

Moreover, this particular form of voluntary, self-governance mechanism (complementing the minimalist approach to specifying technological mechanisms) for regulating the actions the members of inter-networking communities, soon appeared among participants in the regional networks that were forming at the margins of the universities, and even beyond. The creation of many small community networks in U.S. during the early 1980s was propelled by a growing interest in conferencing systems and discussions via mailing lists. Most notable was the UUCP based USENET, a system of newsgroups (bulletin boards) that originally was designed as a forum for UNIX users to discuss their problems and assist one another other with the use of that portable operating system. Very soon USENET grew into a platform for a broad variety of "newsgroups" created by the users, including anti-authoritarian student groups and hacker communities (Hauben and Hauben 1997).

USENET relied on self-organization and also on self-restraint. It was in the USENET context that many rules and norms evolved into what came to be known as the "Netiquette," an informal code of conduct proposed for Internet users. The Netiquette included basic maxims such as "never disturb the flow of information," and "every user has the right to say anything and to ignore anything" – which were viewed as natural extensions of fundamental values of American society, such as freedom of speech and free flow of information. They served at least for a time to extend to a widening circle of inter-networking users a more self-conscious version of the cooperative ethos implicitly subscribed to by the scientific and engineering workgroups who had pioneered in the early development and application of the technology.

Thus, the essential trade-off that was accepted in following the open-architecture's end-to-end design path was to forego assuring full interoperability of applications through engineered standardisation of equipment and network specifications, in order to secure ubiquitous interconnection and freedom of the users to experiment with a variety of applications technologies (both hardware and software). One manifestation of the success of the latter strategy was the emergence of a growing cadre of "inter-networking" engineers that took *de facto* responsibility for achieving and maintaining necessary minimal levels of interoperability through the promulgation of peer-negotiated compatibility standards for the Internet. Starting with the practices of the Network Working Group (NWG) that was loosely organised in 1968 under the leadership of Steve Crocker at UCLA to develop host protocols for the ARPANET, the network distribution of "requests for comments" (RFCs) was initiated in 1969 to facilitate quick dissemination and discussion of ideas and technical specifications by members of what, at that time, was still a small – although geographically dispersed – community of researchers (Abbate 1999: pp. 73-74).

Proposals that seemed interesting were likely to be taken up and tested by someone, and implementations that were found useful soon were copied to similar systems on the network. Everyone who had access to the ARPANET could participate in this process, for although the networks specifications were regarded as military standards ("milspec"), they were not "classified" and therefore remained open and available free of charge. Eventually, as the File Transfer Protocol (FTP) came into use, the RFCs were prepared as on-line files that could by accessed via FTP; today they are readily accessible at various sites on the Web, but the RFC have retained their original formatting, with the same courier font and plain style in which the ARPANET protocol was defined back in 1969.

These were the roots of what has grown into the Internet Engineering Task Force (IETF), the voluntary organisation that was formed in 1986 and is now structured in 100 working groups covering eight to ten functional areas. Only since 1992 has the term "standard" officially been used to describe technical specifications promulgated as having completed the full process of acceptance for the Internet (RFC 1311): two independent implementations must have been shown to work, and to be interoperable. The informal IETF credo, coined by David Clark of MIT conveys the ethos of the Internet's pioneers: "We reject kings, presidents and voting. We believe in rough consensus and running code." (See David and Werle, 2000: p. 19.)

A bedrock of technocratic faith underlies this colourful formulation. For every problem there must be an engineering solution, and optimal solutions to engineering problems will be self-evident to all who are qualified by competence to judge; something cannot be "right" if its adoption has to be authorised by taking a formal vote. That philosophy, and the further legitimation of the rejections of the apparatus of national regulation and international governance that had evolved with the infrastructures of telegraphy and telephony, was given further impetus in the early 1990s by the circumstances under which the NSFNET backbone was opened to commercial traffic and its owner was transferred to the private sector.

(iv) The Privatising of NSFNET and the Rise of the Internet in the 1990's

Although inter-networking originally had been driven by the possibilities of enabling high-speed traffic in massive volumes of digital scientific data and the shared utilisation of the costly mainframe computing capacity needed to process it, recognition of the technology's potential to improve personal communications and collaborative activities quickly began to spread; a widening array of academic and industrial researchers were making use of these networks and the demand for greater capacity in the NSFNET backbone of the system was mounting. The enthusiastic reception of the NSF's promotional efforts saw the number of hosts on the network increased from 213 in August 1981 to 159,000 in October 1989 (see Zakon 2001:statistical annex). Work on ungrading the backbone of the NSFNET to the "T-1 line" level (1.5 Mbps) had been started in 1987 and was completed in July of 1988, and preparations were being made for a further expansion of capicity (accomplished in 1991 when the NSFNET became a T-3 (45 Mbps) backbone (see, CBST, 1994: Appendix A, pp. 237ff. for further details of federal networking programmes in this era).

These and other responsibilities arising from its support of high-speed computer networking were placing increasingly onerous financial and administrative burdens upon the NSF. As a result, under the leadership of Stephen Wolff, the NSF delegated the operation of NSFNET to Advanced Network Services (ANS), a joint venture of IBM, MCI and Merit (see Kesan and Shah, 2001), and from 1989 onwards the Foundation was actively promoting the participation of commercial users in regional networks who could share the fixed costs of expanding capacity, and so might make the infrastracture self-sustaining – at least with regard to the NSF budget. The ensuing problems that arose about the access structure and the arrangements between ANS and commercial ISPs that were taking over responsibilities for operating the network, set the stage for the fourth phase of development.

This, the final phase of NSF's stewardship brought, in remarkably quick succession: (i) a partial easing in 1988-90 of the Authorized Use Policy (AUP) -- which hitherto had proscribed all commercial non-research uses of the network; (ii) an accelerated movement toward privatising the network's core, commencing in 1991 with the addition by ANS of a privately owned backbone; (iii) the creation, and the award in 1994 of private ownership of the new Network Access Points (NAPs) that were needed to connect together the federal networks, commercial backbone networks, and a Very High Speed Backbone Service (vBNS) that was being built as a replacement for NSFNET's research and educational supporting functions; (iv) the migration of regional networks to commercial backbone service providers and the retirement of the NSFNET at the end of April 1995; (v) the withdrawal, announced by NSF in August 1996, of further sponsorship of the four "public NAPs" – which were thereafter to be operated entirely by the private sector – thereby completing the privatising of the Internet (see CSBT, 1999a:pp. 177-179; Kesan and Shah, 2000:pp. 18-26).

Although these matters rarely are discussed in the many accounts of the history of the Internet's development, there are many instructive lessons to be drawn from a detailed examination of the way in which the transition to the private sector control of this communications infrastructure was carried out (see Kesan and Shah (2001) for an extensive treatment). A point that stands out perhaps most starkly in the present connection is that the whole process was intiated and carried through with unusual rapidity and with little if any antecedent history of deliberation about the regulatory framework, and the implications that the terms of the Internet's privatising would have for the competitive structure of the new network services industry that was being launched.

Thus, even though there were plans as early as 1987 that envisaged a future transition of NSFNET to the private sector, little advance work was done in the following four years, so that the organisational aspects of the Internet's emergence may legitimately be cited as another dimension of the story contributing to the "accidental" character of the advent of "the information super-highway." Yet, in this regard, the limited planning of the process appears to have resulting in some outcomes that are decidedly troubling. Kesan and Shah (2001) make a persuasive case that NSF erred initially in entering into a cooperative agreement with ANS that did not envisage the commercial use of NSFNET, and so there were no clear guidelines for NSF when the joint venture that that was acting as the government's sole contractor began to ask for and later sell access to the (publicly funded) NSFNET. NSF's actions in this regard left ANS initially in a position to block the entry of competitors into the provision of backbone services. But, of more lasting consequence, the multiple backbone system – which was intended to create competitive conditions – was designed in a way that contributed to the present highly concentrated structure that has emerged in that market. Furthermore, in designing the NAPs, and in transferring operational responsibility and eventual support to the private sector, the NSF did not put in place performance requirements. The market power of the large ISPs, and the fact that the provision of increased capacity at the public NAPs has been neglected, so that they have become areas for congestion on the Internet, represent specific unhappy legacies of the privatisation process that will bear some further discussion below.

IV. MOUNTING PRESSURES FOR AN ENDING OF "END-to-END"

If the Internet is to be seen as a valuable, albeit challenging technological gift bestowed upon the private sector, this does not warrant regarding it as a static artefact. Quite the contrary, this component of the global information infrastructure continues to undergo technical modifications and the course of this seemingly incremental evolution may well bring about rather radical changes in high-level service performance characteristic of the system.

A number of forces may be seen to have been driving the Internet's technological transformation, but among these perhaps the currently most potent arise from the combination of current perceptions of the "short-comings"and undesireable performance features associated with some widely used applications, and a recognition that new capabilities would have to be added to the Net to meet the technical requirements of a number novel applications that are now contemplated. The critical questions of concern here are whether it is essential to address all of these problems purely through technnological mechanisms, and, where new service requirements necessitate technical "fixes," whether it is possible to implement them in ways that would not entail departures from the end-to-end principles of architectural design.

During the Internet's first two decades the dominant approach among those designing applications was to work within the very tolerant parameters set by the general data transport service provided by the routers in the core of the network. Because the Net was a rapidly expanding facility that could be used by a very wide range of applications, regardless of their individual differences, there was little interest on the part of commercial interests to seek to develop new applications that would, in effect, challenge the existing architectural design principles rather than exploiting them. But, that design regime now has come under increasing pressure from the accumulation of new performance requirements for network services, and the emergence of demands for applications whose features might be more conveniently provide by implementing them "in" the network's core.

Three classes of challenges to the preservation of the Internet's end-to-end architecture have usefully been identified by Clark and Blumenthal (2000): (1) non-cooperative and "untrustworthy" behaviours at the Net's end-points, (2) applications requiring higher levels of service quality in data transport than "best effort" forwarding, and (3) the emergence of large ISPs pursuing new competitive strategies. The following paragraphs indicate the nature of the issues that have come to the fore in these several areas.

(i)Technical responses to non-cooperative behaviours at the end-points

It now is generally acknowledged that the presumption of reliable cooperation among the agents at the end-points, which underlay the approach implemented by the builders of the ARPANET, is completely at variance with the facts of today's Internet. Manifestations of deviations from norms of "trustworthy" behaviour at the end-points of the Net range from the comparatively benign, albeit annoying imposition of email "spam" upon receipients, to unsolicted receipt to messages carrying offensive content, to "denial of service" attacks directed against particular web-servers, and attacks on the functioning of the network as a whole. (For a comprehensive review of the issues involving "trust in cyberspace," see CSTB 1999b.) A variety of technical solutions have been introduced to address these problems. Some confine the remedial mechanism to the applications that run at the endpoints, such as email program filters that automatically delete messages from unrecognized sources, or scan the content for pre-specified "sensitive" text and alert the reader before the message is opened. Other mechanisms, however, are interposed between the sender and the receiver by third parties, and these have the effect of balkanising the Internet by creating enclaves over which discretionary control of information flows can be exercised.

"Balkanisation" is a term that often carried strong pejorative connotations, but, in the context of discussing the architecture of a communications network it may be used in a more neutral way. What usually will determine the value-overtones assigned to the term is the nature of the purposes that are served, and the nature of the purposes that are sacrificed by interposing non-transparent mechanism between users situated on the network's end-points. Another, related taxonomic distinction focuses upon whether the objectives are, or aren't agreed upon by the affected users. On the one hand, the intervening agency may have a controlling role that users are aware of, and to which they assent: business firms or a non-profit organizations, such as universities and private foundations that install "firewalls" are in a position to use these devices to thwart the unauthorized export of particular data files, as well as to block access to specific classes of sites -- whether for the purposes of denying users within the "protected" sphere access to websites identified as providing pornographic material, "hate" publications, or, simply, to time-diverting interactive games, and inessential services that tax the capacity of the organisations servers (as was the experience of some American universities in the Napster episode). In the same way, "filters" are being installed by end-use organisaations and also by ISPs, as they are useful for "traffic analysis" that enables better capacity planning.

The same devices obviously are available for use, and are being deployed by third parties that do not need, and in any case do not ask for the users' consent. According to a recent reports, the government of China has been able in effect to "fire-wall" the entire country, thereby controlling connections with the rest of the Internet in addition to monitoring the content of internally generated traffic (*The Economist*, August 11, 2001, pp.9-10, 18-20) What makes this feasible for an authoritarian government and a business corporation alike is that there are a relatively small number of paths connecting its domain to the rest of the network. Inserting firewalls and filters at those few passage points for traffic is an effective and comparatively low cost means of imposing selective controls on the messages that residents of the domain are able to exchange with the rest of the world; it equally permits the insertion of clandestine traffic analysis and content monitoring by outside parties.

Corresponding key passage points exist in any domestic network, notably where individual users are connected to their ISPs, and at the network access points (NAPs) where the ISPs backbone networks interconnect. Consequently, where ownership of the sub-networks is in private hands, the ISPs are likely to emerge as the vehicles through which even the least authoritarian governments would seek to implement measures of technological monitoring and control in pursuit of public policies (Clark and Blumenthal, 2000: p. 12).

Similarly, given the comparative efficacy of applying legal and regulatory pressures upon large IPS's to impose self-controls over the content they carry, and the continuing drift towards greater and greater concentration in an increasingly trans-national industry, may make it possible for political movements as well as sovereign states to induce the adoption of "voluntary" policies of control.

Conceivably, such labeling policies in conjunction with ISP filtering also might come to be used by governments to restrict, however imperfectly, the carriage of encrypted traffic between sites on the network that were unable to produce certificates of identity accompanied by "authorisations" to engage in exchanges of encrypted data.

(ii) "Enchancements" to meet the requirements of new services

The Internet's TCP protocol, which provides capabilities for reassembling packet in proper order, retransmitting lost packets, and confirming complete delivery, offer a "best effort" quality of service (QOS). Although this has been successful in supporting a wide range of applications running over the Internet, it can not make any guarantees for users as to when, or even whether, data will be delivered. Because voice and video services are degraded by the transmission delays (latency) and delay variations (jitter) that network services like e-mail and Web browsing can tolerate, video- and voice-streaming techniques have been developed that provide a partial enhancement but making use of buffers located on the Net – typically in proximity to the locales where demand for content is most concentrated..

The latter, however, is not a solution that meets the requirements of real-time voice, music and video. Basically there are two approaches to dealing with this problem (see, e.g., CBST, 2001: pp. 98ff). One is to add sufficient bandwidth to keep the packet-queue lengths to the necessary minimum latency levels and, by preventing packets from being dropped when buffers overflow, to suppress jitter as well. The costs of raising the quality of a best effort service in this way for the entire volume of Internet traffic would be astronomical. Attention therefore has been directed to an alternative approach: developing QOS mechanisms that could selectively and less expensively set different service qualities for different (self-identified) classes of traffic, and managing the load placed on the queue so that buffers do not overflow.

Much ingenuity has been devoted to designing a variety of QOS mechanisms. Among those currently under consideration, the least sophisticated, so-called Differentiated Services ("diff-serv" in the specialist's parlance) approach, which that would allow ISPs to provide a quality of service above the default best effort service without permitting the user to have discretion over the particular sessions on the network during which a particular class of service would apply. This will necessitate placing a service class indicator on the packet headers, with the routers of the network being modified to read this label to determine how the packets would be queued; access to any given service class would then be enforced at the edge of the network, using filtering based on IP addresses. A considerably more complex and flexible proposal is referred to as "virtual overlay networks" (VONs). This would call for capabilities to be added to the routers within the network to enable the setting up of virtual networks in which traffic within an individual session-flow would compete with other packets on the same VON, but not with the traffic from other flows (CBST, 2001: pp. 102-103, provides further details and references). There remain unresolved technical issues, with which network engineers continue to wrestle, including how the properties of an the virtual overlay network are to be specified, and how to dynamically administer resources on routers associated with overlay networks. It is perhaps needless to remark that pursuit of such means of "virtually" providing the services of a connection-oriented network most likely would result in a re-engineering of some portions within the network's core, thereby departing far from the Internet's end-to-end design principles.

It is worth closing this brief review by noticing a rather different technical development, which involves the Network Address Translation (NAT) boxes that already are appearing on the Internet, and which have the potential to siginificantly and adversely affect the network's end-to-end performance properties (see Clark and Blumenthal, 2000:pp.11-12). The function of a NAT box is to modify the IP addresses on the packets, and because doing so offers a means of dealing with perceived shortages in Internet addresses, they are typically installed by managers of organisational (intra-) networks, and by ISP's who use them to create "virtual IP addresses" within their domains. As well as also simplifying the management of address space, NATs can help mask the user's identify from other end-points, and they have some other collateral filtering capabilities that leads them often

to be integrated with firewalls. But the deployment of NAT requires adjustments elsewhere, because when the IP addresses of packets entering and leaving a region of the network are rewritten, the box must notify the TCP level where the error checking functions work on the assumption that IP addresses are carried unchanged across the Internet from source to destination. Moreover, some higher level applications protocols also make use of the IP address, which implies that correct operation of those applications will be preserved only if the NAT box understands the design of the application in question. This is a clear departure from the design principle of the Internet, which pushes intelligence of that sort out of the core of the network and confines it to the edges.

But, that is not quite the end of the matter, for, as Clark and Blumenthal (2000) point out, IP addresses now have come to be used in many other ways on the Internet, which also are likely to be disturbed by the operation of NAT boxes. The software site-licensing facilities, for example, have been designed to use the (supposed) stability of the client's IP address as the means of controling the latter's access to the server, and would therefore be likely to reject a user whose ISP, or corporate NAT box presented an apparently different name. These ramifications of the introduction of NAT boxes strikingly illustrate the way in which incremental technological modifications made in a decentralized network facility for benign, but localized purposes, may yield unintended adverse repercussions that extend throughout the system.

(iii) Enabling ISPs' usage-pricing strategies -- from congestion control to rent extraction

Among the early contributions to "the economics of the Internet" perhaps the best known were those concerned with the sources of congestion, and how to deal with them (see, Mackie-Mason and Varian, 1992/1996, 1993/1995a, 1994/1995b). What economists typically brought to this discussion, perhaps all too predictably, was an abstract understanding of the phenomenon of congestion as a negative externality suffered by all users as a consequence of the lack of some effective mechanism restraining individuals' claims on the limited available capacity. Casual analogies were drawn with the phenomena of "over-fishing" and "over-grazing" of common resources, and the spectre was thus raised of the Internet becoming another case of a resource whose utility was seriously degraded by "congestion" arising from the absence of (bandwidth) usage-sensitive pricing. The mantra that subsequently has been imparted to novitiates in the field of "Internet Economics" carries the same message, formulated in a less normative way (e.g., by McKnight and Bailey, 1997: p.12): "Flat rate pricing does not provide an economic congestion control mechanism for bandwidth resource allocation".

Congestion occurs on the Internet whenever the combined traffic needing to be forwarded onto a particular outgoing link exceeds the capacity of that link. The design of the transmission control protocol (TCP) assigns to the sending nodes the responsibility for regulating the flow of packets on the basis of cumulative acknowledgments from (adjacent) receiving nodes of the arrival of packets sent to them. This adaptive control mechanism operates in response to "packet losses" that reach a rate signalling the presence of congestion to the routers that share the link. Thus, when congestion occurs, a packet may be delayed, sitting in an adjacent router's queue awaiting dispatch, and so will arrive later than some other packet from the same message that has not been subject to queuing. The result is delay in the reassembly of all the packets that contain the message, the condition described as "latency" in the language of telecommunications engineers. (When queue lengths vary, and some queues fill up, packets will be dropped by the router and therefore need to be resent, causing variations of the duration of delays and the condition known as "jitter".) Congestion typically is a transient phenomenon, however, lasting only for the interval during which the TCP mechanism adapts to the available capacity by slowing the outgoing packet rate. It can reach drastic levels, however, if the capacities of the nodes available to each router fall below the minimum transmission rate provided by the control protocol.

The mechanism of congestion control provided by TCP, therefore, is simply to push back on the traffic source dynamically, in response to the detection of congestion inside the network, until it no longer is able to accept the offered load. This simple algorithm is incapable of discriminating among the initiators of the offered load, or among various types of applications that are generating traffic. Hence it cannot serve to shape the behaviour of individual users on the Internet, or even that of classes of users. Moreover, this congestion control algorithm is neither enforced on the Internet, nor is it even part of the protocol architecture of some applications that do not implement TCP – such as streaming video and UDP (User Data Protocol). (See CSTB, 1994, p. 189, 201 n. 40.) Those applications consequently can be viewed as taking "unfair advantage" of other applications, such as email that do implement TCP.

Most of the proposals put forward by economists to correct this deficiency have favoured usage-pricing, although their schemes have varied considerably both in the degree of their economic sophistication and their complexity (useful reviews are provided by Cave and Mason, 2001; Gupte, 2001). At the upper end of that scale, the "smart market" mechanism advocated in the pioneering work of Mackie-Mason and Varian (1993/1995a) applies the principles of a "Vickery auction": users would enter bids for network access that indicated a maximum willingness to pay, and routers would recognise the bids attached to each of the data-packets; all packets with bids exceeding some cutoff value would be admitted for forwarding. Given a fixed supply of bandwidth, the cutoff value would therefore be the lowest bid that corresponded to the transmission capacity of the system, and that price would be charged to all users whose bids were accepted. Consistent with marginal cost pricing principles, when there were no bids for network access that fell below the router's cutoff value, the price would fall to zero.

As the authors of this proposal soon acknowledged (Mackie-Mason and Varian, 1994/1995b): "usage-based pricing is itself expensive – it requires an infrastructure to track usage, prepare bills, and collect revenues". A subsequent publication (Mackie-Mason and Varian, 1997) took the matter further by recognising that designing a congestion accounting and billing mechanism for a packet network is not so straightforward a proposition; who should be charged, the sender of packets, or the receiver? Consider the situation in which a user downloads a file from a public archive: both the applications that are "parties to" the communication-transaction originate their own packets, but there is no way for the routers to identify the many packets forwarded from the archive as being responses to the session initiated by the small number of packets carrying the user's request for the file. If such requests resulted in congestion, how could the behaviour of the users be modified by charging the costs to the passive party in the transaction (the archive)? To allocate the congestion costs between the parties, the public archive in this case would have to have installed a billing mechanism, permitting the subsequent reassignment of the charges to the user that had instigated the file transfer.

Just what changes would be required in the architecture and transmission control algorithms to enable the routers to do all this was not considered. But, the design of the Internet's transmission control protocols (TCP) does not allow monitoring the state of congestion everywhere in the network, and so the implementation of the suggested pricing mechanism, like that of quality of service (QOS) schemes, would require monitoring and information collection functions that are not supported and – with the continuing growth of the network – would become increasingly taxing for the simple routers to accomplish in real time. Moreover, the cost allocation and billing requirements for congestion control via QOS systems would call for the collection, transmission and processing of *internal* traffic information, and as well as user bids, and the provision of discretionary network routing capabilities. To imagine all that being implemented without substantial engineering departures from the principles of an "end-to-end" architecture is hard indeed (see Odlyzko, 1998: pp. 26-27; CSTB, 2001: pp. 99-100), and so it seems rather remarkable that the larger implications of such changes have not been more prominent matters of concern for the proponents of such schemes.

More remarkable still is the continuing robustness of the economics literature's fixation on congestion-pricing, the pertinent facts notwithstanding. Congestion was not a major problem on the Internet during the early 1990s, when its opening to commercial traffic first directed attention to the

problem posed by the impending need to introduce usage-pricing; nor has the forecast condition of chronic congestion materialised subsequently. Delays experienced on the Internet will indeed be caused by queues, which are an intrinsic part of congestion control and the sharing of capacity (see CSTB, 2001, pp. 98 ff.). But there can be other sources of delay. Indeed, because ISPs are not required either to collect or release data on transmission delays, dropped packet rates, or other network performance variables, there continues to be much disagreement over the exact extent to which many of the service problems experienced by Internet users are properly attributable to congestion, rather than other causes. The frequently observed delays in the delivery of email, for example, are thought to be almost always the result of mail server faults that result in a large proportion of the load being generated by the re-transmission of packets; and the painful slowness that web-surfers encounter during peak hours is ascribed to non-responding web-servers (see Odlyzko 1998; also, Huitema 1997, as cited by Cave and Mason 2001).

Today congestion generally is understood to be rare within the backbone networks of North American ISPs. The obvious explanation for the failure of chronic, paralysing congestion to materialise under the conditions of "unpriced usage" lies in the rapid expansion of capacity to accommodate the growth of Internet hosts and traffic; and because most of the widely used applications tolerated the congestion control mechanisms provided by TCP. Whether bandwidth increases can continue to keep pace with the growth of demand, of course, depends upon on whether QOS-enabling enhancements are made in the network that will greatly increase the offering of bandwidth-hungry services, and the degree to which competition will either check the ability of ISPs to differentially price such services in a manner that curtails their needs for heavy investment in capacity, or result in rivalries among the larger ISPs to stake out more "real estate" on the Net to attract an expanded customer base.

Instead of appearing ubiquitously throughout the rest of the network, however, congestion does appear to be concentrated at particular "bottlenecks" created by disparities in the provision of capacity. As has been noticed above, the links (exchange points) between ISPs – and especially the public NAPs – are as a rule much more heavily congested than the links *within* the service providers' respective networks (See, e.g., Odlyzko 1998, CSTB 2001: pp. 99, 117.). Although the links between customers' local area networks (LANs) or residences and their ISPs are also frequently congested, the difficulty arises from the organisational delays or the expense entailed in increasing the capacity of the connection. Persistent congestion has been documented at several international links, where long and variable queuing delays, as well as high packet loss rates, have been measured. (See Paxson, 1999, and the discussion in CSTB 2001: pp.99-100.) Here again, however, the proximate source of the problem appears to be rooted in institutional circumstances affecting the provision and allocation of capacity at strategic connection points, rather than the endemic condition of unrestrained bandwidth usage envisaged by economic theorising.

A cynical commentator might conclude that the stream of ingenious proposals from economists to fix the problem of congestion on the Internet, in typically ignoring the possible strategic explanations for congestion at the public NAPs, and proposing the introduction into the network's core of the intelligence needed to operate a sophisticated pricing mechanism, come down to the expedient of making the network less and less like the Internet, and more closely akin to a connection-oriented conventional PSTN. Quite obviously, however, had such a design been embraced to begin with, many other difficulties posed by the peculiar open-architecture would have been obviated as well. Along with removal of the obstacles to a mass transfer fee-for-service business models, this would reduce the myriad *practical* difficulties that local communities linked to the Internet now encounter in seeking to control the content of messages bearing "objectionable content." In a connection-oriented system it is much more feasible to rapidly and accurately identify the locations, if not the identities of agents engaging in the electronic transmission of content which recipients deem to be pernicious -- and to set about mobilising political, social and legal counter-measures. There would, therefore, be less need than presently exits to devote resources to the development of the still rather coarse-grain "geo-locator technologies" that now are being use to create targets for "direct mail" advertising and sales techniques based upon the characteristics of the receipients' neighorhood; or to figure out whether such technologies can be made sufficiently reliable to be employed to control the distribution

of objectional content on the Internet, in the ways that would parallel the familiar content-regulating actions of political authorities who can identify the originating parties and have legal juridication over geographical territories in which they are situated.

Whether or not the removal of anonymity, and the re-imposition of greater controls on individuals' access to content are desirable in some circumstances, is quite another matter (see Engel and Keller, 2000). The point is simply that the congestion-pricing solution envisaged for the Internet is not the narrow matter of economic efficiency that economists have appeared to be presenting; its implementation would require an architectural revolution in which the Internet as we know it would have disappeared. Correspondingly, in that "brave new world," debates about the conflicting desiderata of privacy, anonymity and security would continue, but they would cease to be policy matters that had a peculiar "Internet" aspect and would simply reprise the issues that society has found ways of resolving for other communications media --- physical newspapers and books, plain old telephones, radio, movies, and TV (see de Sola Pool, 1990).

Will the commercial pressures to insert new capabilities into the core of the network really have the deleterious effects envisaged, and if undesireable consquences materialised, would it not be possible to restore the status quo ante? Yet, the likelihood is that even the unintended ending of an integral "end-to-end" Internet would not be readily reversible, and that the benefits thereby lost might prove difficult if not virtually impossible to recover on a later, improved successor to the global information infrastructure.

This last point deserves further elaboration, which can conveniently be provided by returning to consider the concrete issue of permitting cable companies in the ISP market first to create proprietary sub-networks on which QOS technologies are used to offer differentiated service choices to subscribers. Users of a particular service, however, would have access only to the music and the video that their ISP had designated, possibly also to a designated IP voice telephony service, and might be similarly "locked in" to a particular suite of other Web-based services and applications software. Once that structure was in place, however, the ISP in question might well be receptive to allowing compatibility between this sub-network and other, similar sub-networks. The economic logic of this situation differs from that which governs in the general analysis of compatibility standardisation for network interoperability – where it is generally found that small networks seek connectivity with larger ones, and the latter have stronger incentives to remain aloof from rivals of comparable size. (See David and Greenstein, 1990, on the research literature; Shapiro and Varian, 1999: esp., chs. 7, 9 on strategies in "standards wars".) By linking with similarly sized networks, an ISP with a large network base could offer subscribers other enhanced services that are latency-sensitive, such as voice telephony, and a larger choice among the set of pre-selected applications. The value of integrating to achieve compatibility with smaller ISPs would remain comparatively small, and so, in this market setting, the dynamics lead towards a high degree of market power concentrated in the hands of a small number of ISPs, and a large fringe of ISPs whose clientele remains cut off from these enhanced services.

Thus entrenched, the dominant ISPs would be in a position to extract some if not most of the rent that might otherwise flow to the developers of applications innovations, in exchange for making these available for use by their clientele. Lacking that access, the developers would be confined to exploiting niche markets at the fringes of the network, where their products would remain beyond the reach of the subscribers to the large ISPs. Nothing in this picture suggests that the emergent structure of a partitioned network would be likely to be voluntarily dismantled by the incumbent, vertically integrated ISPs, nor successfully attacked by an entrant possessing a novel and superior application technology. An entrant with the capital resources required to establish a new, competitive vertically integrated ISP, moreover, would have every incentive to seek compatibility with an existing large service provider and, were the newcomer aggressive might expand by stealing the original incumbent's clientele. But, in addition to requiring the financial backing to create the additional network capacity required for the implementation of that strategy, the successful entrant would replicate the initial situation, and pose an even greater entry barrier to the next innovator.

A mitigating consideration to be noted in the foregoing connection is that although the foregoing technological "enhancements" of the Internet would create new opportunities for ISPs to extract greater "rents" (consumer surplus) from their customers by means of discriminatory pricing schemes, the strategy of vertical bundling of networking services and Internet-based applications nevertheless would provide additional benefits for a large segment of the Internet population. The technologists who created an end-to-end architecture, and who value it especially for the support it provided to applications innovators, are less burdened than the typical Internet user by having to install, configure, upgrade and maintain the software of each and every one of the rapidly growing number of applications that must be attached at the networks' end-points. This state of affairs can be expected only to become more burdensome. As Clark and Blumenthal (2000: p.4) perceptively observe:

> "The importance of ease of use will only grow with the changing nature of consumer computing. The computing world today includes more than PCs. It has embedded processors, portable user-interface devices such as computing appliances or personal digital assistants (PDAs, such as Palm devices), Web-enabled television and advanced set-top boxes, new kinds of cell-phones, and so on. If the consumer is required to set up and configure separately each networked device he [sic!]owns, what is the chance that at least one of them will be configured incorrectly. That risk would be lower with delegation of configuration, protection, and control to a common point, which can act as an agent for a pool of devices. This common point would become a part of the application execution context….there would no longer be a single indivisible end-point where the application runs."

While pointing to the treat to the preservation of the open-network architecture, this acknowledges that the creation by ISPs of enclaves containing advanced services would be one way in which the multitude of less technically sophisticated users could obtain specialised (and correspondingly standardised) network applications-integrating services. Thus, in regard to this issue as is the case in so many others, network policy-makers face the classic "trade-off" of securing the immediate benefits of standardisation by sacrificing the technological flexibility that is conducive to future radical innovations (David 1995).

V. POLICY PRIORITIES AND PROTECTION OF THE INTERNET'S ARCHITECTURE

It has been seen that among the many technological "fixes" proposed for enhancing the Internet's performance, some are not so innocuous because they would entail inserting intelligence into the "core of the network". The likely impact of these induced innovations therefore would be the alteration of the distinctive "end-to-end" architecture, pushing the future path of the network's evolution more towards emulating the performance features (both good and bad) associated with a "connection-oriented" telecommunications system – the familiar paradigm of which exists in the public switched telephone networks. (See David and Steinmueller, 1996 on this prospect.) Will the changing balance among the interests of the communities using the information infrastructure, inevitably, force a sacrifice of the global infrastructure's transparency and openness, thereby raising new barriers to the invention and diffusion of valuable applications? Inasmuch as a technological drift away from the original Internet's end-to-end architectural design should not be regarded as an inexorable process lying beyond the reach of social control, there is scope for policy interventions to check such a course of evolution. It must be hoped, then, that promoting wider understanding of the issues that are at stake can increase the political feasibility of arriving at rational policy priorities. At least, that is the spirit in which the following commentary on the identification and balancing among confliciting "goods" will proceed. .

(i) QOS "enhancement" of the broadband Internet – a matter of benefits and costs

A first appropriate step is to ask whether the net impact of any proposed movement in that direction would be socially beneficial. In view of the prospective emergence of a broadband Internet on which QOS will be more widely implemented by ISPs competing for customers while seeking the means to charge what the (multimedia) traffic will bear, the question might be asked whether the time has come for "end-to-end" to end. I could be argued that inasmuch as the days of "Internet1" as a unified global infrastructure providing a receptive platform for rapid innovation and experimentation with networks are numbered, the best course of action would be make whatever changes are required in the core of the network to quickly reap the benefits of the available new services on a "users' Internet." That is to say, we should come to terms with the outcome of the evolutionary dynamics driven by the needs of the maturing market for a differentiated internet service, and think about other ways to provide a network environment that would stimulate the continuation of "amazing innovations."

Such a view would counsel turning attention to the construction of a separate, very high speed inter-network as the test-bed and experimental commercial market for advanced services, which would be designed to provide the features of openess and flexibility that have proved so encouraging to the development of more powerful digital technologies. This might be called *Internet2+* to distinguish it from the actual federally funded backbone created to continue NSFNET's research role. There is something to be said for this vision of a cyclical regeneration of a new inter-networking environment that would revive some characteristics of the original. It acknowledges the important symbiotic relationship between the mature PSTN infrastructure upon which packet switching and the novel technologies of the ARPANET and NSFNET could be erected; and it recognizes the fertility of experimental research communities as sources of "user-designed" technological innovations. But, unfortunately, it ignores the crucial fact that an important function aspect of the historical experience cannot be replicated or revived by these means.

The nub of the problem is that to develop innovations that are readily available for deployment on the Internet as it exists, one needs a test-bed with its technical features. Yet, for the communities that would have access to *Internet2+*, and especially for those groups that are engaged in advancing the frontiers of network engineering, the high value use would be to develop applications that utilised the enhanced properties of that infrastructure rather than the more limited capabilities of Internet1 -- or the still less accommodating infrastructure into which the latter would be tending to evolve. To make this observation more concrete, it may be noted that in May 2000 Ipv6 was implemented on the actual high-speed Internet backbone network known as Internet2 (see Zakon, 2001). Since the 128-bit address code specified by the IP6 protocol vastly increases the number of available domain addresses, permitting the assignment of unique addresses to individual microprocessor controlled devices that can be reached by any telecommunication channel, much of the theoretical and practical engineering challenge will be to develop ways of controlling such devices, and integrating their functioning into larger and more complex systems. Protocols for the remote management of household appliances such as coffee-pots and electric toasters, will soon be multiplying among the RFCs emanating from the Internet Society (formerly the IETF, see, e.g., RFC 2324: "Hyper Text Coffee Pot Control Protocol HTCPCP/1.0," 1998). This and related trajectories of research and invention, however, will necessarily become increasingly disconnected from operating conditions on the broadband network that is already deployed -- unless the problems of governing Internet1 can be successfully resolved.

Yet, at present nobody knows how or whether it would be possible to stop that from happening by "migrating" the mass of Internet users from the IPv4 protocol that has become universally deployed,) to IPv6 (CSTB 2001:pp.77-81). With the possible exception of the planned use of IPv6 by suppliers of so-called "third generation" wireless devices that are being developed to succeed mobile telephone systems, equipment vendors and service providers are not offering the new protocol – presumably because few users presently appear to see sufficient advantage in the change to make the pain and cost worthwhile for them. This is the usual "chicken-and-egg" problem that can cause coordination failures in market-guided standardisation processes, which is well is well-recognised by the economics literature on network externalities and interoperability standards (see, e.g., Farrell and Saloner 1986, David and Greenstein 1990, Shapiro and Varian 1999). Unfortunately, in the present situation of the Internet, there no longer is the alternative mechanism of a central, governing agency

or authority with the power to compel a coordinated switch of the sort that occurred when Defense Advanced Research Projects Agency (DARPA, formerly ARPA) mandated the transition from NCP to TCP on 1st January 1983, and again, when NSF's "diktat" obliged users wishing to connection with the NSFNET to install the TCP/IP protocol stack

(ii) Upgrading for Internet telephony – is this trip necessary?

Would the enhancements in the quality of differentiated services, and in the ability of service providers to engage in price discrimination among the users of the Internet, compensate for whatever losses might be entailed in terms of curtailed future scalability and a slowed pace of innovation in applications? Several grounds for scepticism regarding the value of the gains seem worth keeping in mind.

To begin with, the incremental social benefit of upgrading the Internet to carry real-time audio traffic is not obviously overwhelming, given the existence of other technological means of providing a large part of the world's population with access to voice telephony (via cellular radio and satellite transmission) at lower fixed costs than those entailed in laying copper wire or fibre-optic cabling. To be sure, Internet telephony could be integrated into new, multi-media services. Yet, there is a disjunction here between a strategy directed toward opening profit opportunities in the developed economies -- to elicit continued private sector investment in augmenting the broadband infrastructure available to users in those countries -- and a policy that also takes account of the situation in the world at large.

While cell-phone technology has opened the benefits of rapid, global communications to large cohorts in the developing economies, it remains unsuitable for sparsely populated regions and geographically remote sites, just as it is not capable of supporting the very high bandwidth communications that are likely eventually to be in demand there. But systems of low earth orbit satellites (LEOS), which are designed to provide two-way, low-latency, point-to-point transmissions, will be available to fill these significant service gaps. According to expert engineering opinion, the seamless linking of LEO satellite constellations into the world-wide communications infrastructure is a development that can be expected to take place in the relatively near future. (Private communication from Robert Spinrad, 9 May 2000.)

For the developing economies, however, it is accepted that even to provide substantial narrowband coverage, considerable amounts of public funding for upgrading existing telecommunications infrastructures would be necessary; and some of that is likely to be provided by subsidised loans and transfers through multinational cooperative agencies. It must therefore be recognised that the social rate of return on public (and private) investments in this infrastructure would be reduced substantially if the present core of the Internet were to be modified by engineering changes that deviated from the principles of "end-to-end". To permit alterations of the architecture of the backbone networks in the high income countries, in order to provide users there with Internet voice telephony (along with business or entertainment services integrating real-time video), would effectively mean curtailing the access afforded newly connected users in the world's poorer societies to existing information tools and global data resources.

Thirdly, the claim that (with further upgrading of the technology to permit differential pricing of services) it would be possible to eschew the regulation of business activities on the Internet seems an illusory hope in the case of the high-income economies. The drive towards introducing QOS to enable new services such as real-time video and audio – even if voice telephony via the Internet were not considered a sufficiently compelling goal – has the already noticed potential to result in vertical bundling of applications and other services by ISPs. A consequent reinforcement of existing trends towards greater consolidation and concentration in that market, would most likely issue in government interventions to preserve competition in Internet services that presently remain

unregulated. Such pressures already exist for legislation or executive action to end the asymmetry between US policy towards voice telephony on the (long-regulated) PSTN, and (currently unregulated) Internet telephony. (See CSTB 2001: pp. 27, 170-175, for concerns that the regulatory framework for the PSTN might thus be imposed, inappropriately upon the Internet's very different technology and industry organisation.)

(iii) Monitoring and filtering content on the Internet

Turning then to the sources of pressure to address the problematic technical aspects of the Internet, it has been noted that the incremental value of the contemplated "technology fixes" may turn out to be considerably smaller than might now appear to be the case. This is especially so in regard to dealing with the problems created by "socially or culturally objectionable content" or "communications for legally or politically unacceptable purposes." The reason is simply that when it comes to discerning the likely nature of contact by inferences based upon traffic analysis, or use of "geo-locators" inserted at key points in the network, each advance in technique to date seems to have elicited corresponding advance in the counter-measures. (see, e.g., CBST 1999b; Spinrad, private communication). For example, a simple strategy now available for users to achieve anonymity and protect their communications from observation by third parties -- whether private or governmental -- is to route traffic through their own third parties, and thereby remove identification in the messages. Moreover, various services have become available to prevent traffic analysis, and there now are more sophisticated anonymous mail relay techniques, such as the "nym server" which allows users to construct a route for messages on which the routing details are hidden by encryption from the ISP, and other third parties (Clark and Blumenthal, 2000).

The foregoing remarks address possible discrepancies between the private incentives driving the Internet's technological evolution, and the social value of the enhancements that would thus be achieved. They have not touched on the need to explore engineering improvements that can be implemented (at the edges of the network) in ways that would not compromise performance attributes that derived from the Internet's end-to-end architecture. An example of the latter, content labeling conventions, whose use of which might either be voluntary or enforced upon some content providers, would enhance the efficiency of filtering at the end-points of the network.

But an another important set of alternatives to introducing control mechanisms in the network's core, which remains to be consider is the large class of *non*-technological options. In view of the fact that the origins of many of the vexing dysfunctionalities of the Internet derive from the historical displacement of the technology system from the peculiar, highly regulated behavioural and organisational contexts within which it was created and initially used, an obvious option to be considered is the restoration of some of the former modes of regulating users' behaviours. The Internet may have been a technology that quite by accident was well-attuned to the *laissez-faire* spirit of the era in which it was publicly introduced. Yet, an ideologically driven commitment to go on thinking exclusively in the same vein about ways to overcome the problems posed by the "network of networks", rejecting "social engineering" in favour of solutions found through "Internet re-engineering", is most likely to sacrifice the Internet's unique and valuable pro-innovation features. There is no a priori reason to conclude that the most efficient solution path is one that relies solely upon "fixes" that can be technologically implemented. Proposed regulation and interventions by public authorities continue to be opposed on the argument that such actions are inimical to the Internet's survival as a global interaction space free from governmentally imposed structures of *social* regulation.

Current rhetorical support for relying upon engineers to fix whatever might really need mending, rather than letting legislators and lawyers loose in cyberspace, presents a curious mixture of attitudes. These are compounded from the libertarian philosophy that is pervasive among survivors of the Internet's pioneering user-groups, strains of anarcho-syndicalism that have emerged in the ethos of the latter-day "hacker culture", and the generic *laissez-faire* disposition of the Internet's more recently arrived community of business entrepreneurs. The holders of pro-commercial and anti-commercial

sentiment alike appear quite comfortable making common cause against the "intrusion of government regulations" that are *socially* engineered. This, it should be recognised, presents an essential political and philosophical position, quite distinct from the utilitarian rationale that would give priority to preserving the distinctive end-to-end architecture of the Internet – especially inasmuch as serving the latter priority might call the development of new, institutional mechanisms of governance.

Lawyers looking at the evolving Internet are naturally disposed to pose this issue in terms of a political choice between the regulation of human actions by laws or governance by "*Code*" – the encompassing term used by Lessig (1999) in referring to the architectural configuration of networks and the location of access points, the design of hardware, operating systems, languages, data formats and applications software. Economists, it would seem, would have something helpful to contribute to debates on these questions, by directing attention to the relative costs of alternative modes of regulation in network environments, especially in view of the significant externalities and irreversibilities that are likely to be entailed by introducing either technological or institutional modifications in the existing regime. Furthermore, approach some questions that involve the governance of human behaviour in cyberspace from the perspective of the "economics of crime and punishment" may also be a useful way to mediate in debates between the engineers and the lawyers: the quest for perfect technological mechanisms of detection and suppression of malefactors is only relevant in a perfect world, and it is possible to compensate for reduced probabilities of being caught by raising the penalties visited upon those who are. This approach may not be good enough in some areas of concern, and other technological safeguards will be needed to protect humans and vital technological systems alike from grave damage. But much of the law afforded under the law has been found to work tolerably well with this "mixed" approach.

For those reasons and still others, the relevant policy questions ought not to be construed in terms of making "either or choices." It is important to resist the rhetoric of much contemporary discussion of economic policy, which tends to offer only extreme alternatives. Participants are too often driven into opposing camps, one side calling for the introduction of government controls, and the other placing its faith upon the further development of decentralised, automatic, supposedly neutral and ("market-like") regulatory mechanisms that can better resist political manipulation and so preserve greater scope for human volition. The following statement exemplifies the polarising impact of applying the "technologists' Internet philosophy" to decide upon the best means of protecting privacy on the Net:

> "[t]he cyperpunk credo can be roughly paraphrased as 'privacy through technology, not through legislation.' If we can guarantee privacy protection through the laws of mathematics rather than the laws of men and whims of bureaucrats, then we will have made an important contribution to society. It is this vision which guides and motivates our approach to Internet privacy." (Goldberg et al., 1997, quoted in Clark and Blumenthal, 2000, p.28, n.52.)

A full-blown systems design approach, by contrast, would hold that if the benefits of the Internet's end-to-end architecture are to be retained, some technological solutions simply cannot be substituted for other, socio-legal modes of governing the behaviour of agents on the Internet. Rather than being viewed as antithetical substitutes, the potential complementary of technological and institutional mechanisms governing the digital communications infrastructure need to be explored in a coordinated manner.

There is thus a case to be made for devoting greater attention to matching the technological innovations of the Internet by mobilising other, *non-technologically implemented* modes of regulation. Greater consideration surely is worth directing to the design of legal, political and social rule structures and administrative procedures, of the kind that proved to be efficacious in supporting successful economic exploitation of previous technical advances in communications networks. In this connection it is worth recalling that the oldest international treaty organisation in existence today is the International Telecommunications Union (ITU). This institution, which began its life in 1865 as the International Telegraph Union, provided the model in whose image virtually all subsequent

international treaty-based organisations were created (see David and Schurmer 1996; Schmidt and Werle, 1998). While that may suffice to suggest the possibility that fruitful innovations in international rule-making *fora* can be driven by the opportunities, or problems, that new technologies create, there is no doubt that today very formidable challenges are posed for the adaptive co-evolution of international laws governing cyberspace (see, e.g., Gamble, 1999.)

Globalisation and global information networks call for the internationalisation of rule-making, both in terms of de-regulation and re-regulation. This applies to issues of network access as well as the regulation of content (see, e.g., Grewlich, 1999). Traditional civil and criminal law notions of legal liability for the provision of harmful content are challenged by the provision of information and entertainment over global networks; presently there is no international agreement as to which of the actors on the multimedia chain (network operator, ISP, content packager, or content producer) should be held liable for content deemed "harmful" or injurious, by local or national community standards. Nor is the notion of common international criteria for the setting admissible "local standards" at all well defined. A similarly difficult challenge pertains in regard to issues of jurisdiction – i.e., determining which state should have the legal right to intervene in the operations of networks over which content is being disseminated globally.

National governments belonging to the EU lately are being forced by such questions to review public law traditions, and to seek innovative forms of regulatory cooperation. The existence of new common European legal and regulatory standards, and the concerns of national entities for cross-border effects arising within the unified EU market from persisting national differences in administrative and legal regulations, obviously, are powerful forces impelling these harmonisation efforts. But, whether the development of regional agreements on trade liberalisation covering other parts of the world would alone be sufficient to provide greater impetus for cooperative approaches in forming new international governance mechanisms, remains far from clear. It certainly constitutes a question deserving further examination. (See the forthcoming CSTB Committee Report on *Global Networks and Local Values*, esp., chs.5, 9.)

VI. CONCLUSION: TOWARDS A MORE POLICY-RELEVANT "INTERNET ECONOMICS"

Even as the Internet "comes of age", the technology of the global information infrastructure and the organisation of the communication service industries based upon it continue to undergo significant changes. The main message carried by the foregoing discussion is that many engineering proposals and policy recommendations that have been presented as incremental modifications to enhance the performance capabilities of the Internet actually may have radical implications for the future course of its technological evolution. These have been seen to involve rather esoteric matters that might appear best left to be decided by engineering specialists, and experts in the intricacies of telecommunications regulations. But decisions taken in those realms will powerfully shape the future performance characteristics of the Internet. In that way, they will have important consequences for the nature, size and distribution of the economic and social benefits that it yields.

Bertrand Russell once remarked that we must "tolerate specialists because they do good work". Perhaps it would be more generous to speak of "appreciation" rather than toleration, but the point remains that in matters whose potential implications for human welfare are as important as those at hand, more than narrow expertise is wanted. The story of the Internet's development justly can be presented as a remarkable case of "success by design" (CSTB 2001, p. 34 invokes this phase in discussing architectural principles). Equally, it may be read as a path-dependent tale of fortuitous engineering design decisions that were made with little consideration for aspects that have turned out to be problematic for many of the purposes, and social contexts in which the resultant, wonderfully open and flexible technology would be used. (On concepts of irreversibility, path-dependence and "path-constrained melioration", see David 2001.) As societies around the world continue to wrestle with difficult technical challenges and policy quandaries that have their origins in historically remote

decisions that proved to be essentially irreversible, an obvious question to be asked is whether it has become possible now to proceed differently.

That is to say, is it possible to arrive at policy commitments affecting the future technological evolution of the Internet (and its successor networks) that are any more fully informed and forward-looking than those of the past, or do we need to press forward as before and trust largely to a mixture of good luck and future ingenuity in coping with the outcome? The response here has been to try to move in the former direction, however modestly. Thus, the discussion has sought to identify both the range of often conflicting concerns, and the array of often mutually compatible policy instruments, that need to be kept in view during what should be inter-connected processes of debate and decision-making in the areas of telecommunications engineering, institutional design, and the development and application of judicial and administrative law.

The central policy issue can be structured roughly in the following terms. The end-to-end architectural design of the Internet can be regarded as "a public good" that facilitates certain (publicly enjoyed) benefits, e.g., scalability, extension of connectivity at low incremental costs, ease of development and wide deployment of new applications – which stimulates innovation and the benefits that users derive therefrom. Therefore, there are conflicts with other "public goods," such as trustworthiness of communications (security, privacy), protection from terrorist actions coordinated through the Internet, ease of "setup" for the mass of users. In addition there are conflicts with "private goods," such as freedom from spam" and offensive content, that can be obtained at lower private expense by technological "fixes" that would be introduced into the core of the network.

Three broad guidelines then emerge for the formulation of regulatory approaches. First, where the conflict is with "private goods," the principle of protecting the end-to-end architecture generally should dominate, even if confining the implementation of technical fixes to the edges of the network are less efficient from an engineering perspective and impose higher private costs upon users. Second, where the provision of the "private good" only is feasible by altering the open, transparent character of the network core, arguments for adhering to the "end-to-end design principle" have no real force: the trade-off becomes not one of "how" but of "whether" to do it. The loss of the "public goods" benefits should then be weighed against the private gains of the new functionality that the engineering changes would provide for users. Third, some still more difficult situations must be acknowledged in which there will be conflicts among different forms of "public goods" on the Internet. Consideration of the distributional issues this is likely to raise is unavoidable, so that the questions for policy-making are inherently political; to opt simply to implement the technological modifications that can be implemented as least direct cost would not only ignore this, but generally will fail to guarantee a solution that was optimal even when viewed from the narrow perspective of economic efficiency. This set of "guidelines" is not incongruent with the positions recently articulated in Clark and Blumenthal's (2000) reconsideration of how "end-to-end principles" should be applied in policy debates concerned with regulation of the Internet.

The "historical economics" approach (David, 2001) that informs much of the foregoing discussion carries some additional, and potentially more provocative suggestions for rethinking the economics of the telecommunications regulation in the age of the Internet. Because economic analysis of industrial organisation and public regulation of telecommunications utilities was developed with reference to industries based upon a *mature* network technology, practitioners in this area remain too inclined to start from the assumption that "the technology is given". This is seldom true, and it is palpably misleading when applied to the situation of the Internet. The first lesson, then, is to think in terms of the way in which the structure of the existing markets, and the uneven and uncoordinated regime of regulation and non-regulation, induces research and technological innovation to take some directions, while discouraging it from proceeding in others.

There are two additional, corollary suggestions for economists who rightly think they should have something to contribute to solving problems in the allocation of resources on the Internet. First, from the view of the technological configuration of the Internet as an endogenous, rapidly co-evolving state

variable there follows the simple principle that economists should start by understanding the features that already are distinctive but subject to transformation. For quite understandable reasons, such as the lower average cost of prescribing remedies for problems you have treated many times before, excessive attention has been devoted to the topic of network congestion and pricing remedies to control it. Indeed, this emerged as the paradigmatic subject matter of mainline "Internet Economics". It has taken some while for the designers of sophisticated usage-pricing mechanisms (and some QOS implementations) to come to a realisation that implementation of their proposals would require substantial re-engineering inside the network, modifications that apart from their immediate cost would in effect sacrifice the benefits deriving from the end-to-end principles of Internet architecture.

This, of course, is not the only instance of a misdirected economic policy recommendation, and it will certainly not be the last. Its significance here is to be found in the fact that it stems from the more general practice of casually transferring to the sphere of "internet economics" analyses and policy prescription that had been developed in the context of mature telecommunications networks (i.e., the PSTN). Similarly, a good bit of prominence has been given to the discussion of principles that should govern optimal pricing of access to the transport/bearer layer of the Internet, a matter of undoubted importance for existing and would-be service providers. In a technologically dynamic network setting such as that of the Internet, however, the feasibility and terms of entry also depend on "non-price" policies, including those affecting technical compatibility standards, and regulations governing the interconnection strategies of incumbent service providers (see Cave and Mason, 2001). Over the long-run, the "technical rules of the game" affecting physical interconnection are likely to be more consequential than pricing formulae in their effects upon the growth and distribution of available bandwidth, competition in the ISP market, and the rate of innovation in applications on the Internet.

In the current, fluid state of the technology, it seems imperative for economists to work more closely with members of other disciplines in assessing the societal implications of specific proposals to modify the technology and governance institutions of the Internet. A foreseeable and highly desirable outcome of embarking upon such a program would be the transformation of "Internet economics" into a more policy-relevant area of inquiry; a sub-discipline that was effectively defined by its recognition of the distinctive technical constraints and potentialities of the existing technology, as well as by the regulatory issues posed for the Internet industry by its co-existence with industries and institutions that originally were formed on the basis of quite different communication facilities. Members of other well-established disciplines – principally, specialists belonging to the telecommunications engineering community, and lawyers versed in regulation and international law – who would be drawn into this process may be just as discomforted as economists in having to yield some sovereignty in their accustomed domains of expertise. Nevertheless, the price to society of indulging each among the needed assembly of research communities in the comforts of professional autarchy now seems too high to continue to be easily borne.

REFERENCES

Abbate, Janet Ellen (1999), *Inventing the Internet*, Cambridge MA: The MIT Press.

Baran, P. (1964), "On Distributed Communications Networks," *IEEE Transactions on Communication Systems*, March..

Berners-Lee, Tim with Mark Fischetti (1999), *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*, San Franisco, CA: HarperCollins.

Cave, Martin and Robin Mason (2001), "The Economics and Regulation of the Internet," *Oxford Review of Economic Policy*, 17(2), forthcoming in September, 000-000.

Cerf, V. G. and Kahn, R. E. (1974), "A Protocol for Packet Network Interconnection," *IEEE Transactions on Communication Technology*, vol. COM-22 V5, (May), 627-641.

CSTB (2001), Computer Science and Telecommunications Board, National Research Council, *The Internet's Coming of Age*, Washington, D.C.: National Academy Press.

CSTB (1999a), Computer Science and Telecommunications Board, National Research Council, *Funding a Revolution: Government Support for Computing Research*, Washington, D.C.: National Academy Press.

CSTB (1999b), Computer Science and Telecommunications Board, National Research Council, *Trust in Cyberspace*, Washington, D.C.: National Academy Press.

CSTB (1997), Computer Science and Telecommunications Board, National Research Council, *The Evolution of Untethered Communications*, Washington, D.C.: National Academy Press.

CSTB (1994), Computer Science and Telecommunications Board, National Research Council, *Realizing the Information Future*, Washington, D.C.: National Academy Press.

Cerf, V. G. (1997), "A Brief History of the Internet and Related Networks," The Internet Society, available at: www.isoc.org/internet/history/cerf.html.

Clark, D. D. and Blumenthal, M. (2000), "Rethinking and Design of the Internet: The End to End Argument vs. the Brave New World," Working Paper, MIT Lab for Computer Science.

David, P. A., (2001), "Path Dependence, Its Critics and the Quest for 'Historical Economics,'" *Evolution and Path Dependence in Economic Ideas: Past and Present*, edited by P. Garrouste and S. Ionnides, Cheltenham, England: Edward. Elgar Publishing.

David, P. A. (1995), "Standardization Policies for Network Technologies: The Flux Between Freedom and Order Revisited," ch. 3 in *Standards, Innovation and Competitiveness: The Political Economy of Standards in Natural and Technological Environments*, (R. Hawkins, R. Mansell and J. Skea, eds.), London: E. Elgar, 15-35.

David, P. A., and Greenstein, S. (1990), "The Economics of Compatibility Standards: A Review of Recent Research, *Economics of Innovation and New Technology*, 1(1&2), 3-41.

David, P. A., and Shurmer, M. (1996), "Formal Standards-Setting for Global Telecommunication and Information Services," (with M. Shurmer), *Telecommunications Policy*, vol. 20 (10), December, 789-815.

David, P. A. and Steinmueller, W. E. (1996), "Standards, Trade and Competition in the Emerging Global Information Infrastructure Environment," *Telecommunications Policy* , vol. 20 (10), December 1996, pp. 817-830.

David, P A. and Werle, R. (2000), "The Evolution of Global Networks: Technical, Organisational and Cultural Dimensions," [Unpublished background Paper for DAAK-CSTB Committee on Global Networks and Local Values]. Stanford Department of Economics, and Max Planck Institute for the Study of Societies, March.

de Sola Pool, I. (1990), *Technologies Without Boundaries: On Telecommunications in a Global Age*, Cambridge, MA and London: Harvard University Press.

Engel, C., and K.H. Keller (eds.) (2000), *Understanding the Impact of Global Networks on Local Social, Political and Cultural Values*, Baden-Baden: Nomos Verlagsgesellschaft.

Farrell, J., and Saloner, Garth (1986), "Installed Base and Compatiblity: Innovation, Product Preannouncements and Predation, *American Economic Review*, 76(4), 940-955.

Goldberg, I., Wagner, D. and Brewer, E. (1997), "Privacy-enhancing Technologies for the Internet," available at: www.cs.berkeley.edu/~daw/privacy-compcon97-222/privacy-html.html.

Gamble, J. K. (1999), "New Information Technologies and the Sources of International Law: Convergence, Divergence, Obsolescence and/or Transformation," *German Yearbook of International Law*, Berlin: Duncker & Humbolt, 170-205.

Grewlich, K. W. (1999), "Access to Global Networks—European Telecommunications Law and Policy," in *German Yearbook of International Law, vol. 41--1998* , Berlin: Duncker & Humbolt, 9-54.

Gupte, R. P. (2001), "Pricing to Control Congestion: An Economist's Bias," Trinity Term Research Paper in Economics 168X, Stanford University Centre in Oxford, June.

Hafner, Katie and Lyon, Matthew (1996), *Where Wizards Stay Up Late. The Origins of the Internet*, New York: Simon & Schuster.

Hauben, Michael and Ronda Hauben (1997), *Netizens. On the History and Impact of Usenet and the Internet*, Los Alamitos CA: IEEE Computer Society Press.

Hameri, A., and Norberg, M. (1998), "Creating a Solution for Document Sharing- The Early Years of the WWW," *Journal of Product Innovation Management*, 15 (2), 45-61.

Headley, W. (1995), "The Organizational Dynamics of Computer Inter-Networking: Insights from a History of BARRNET," Working Paper in the History of Philosophy of Science Program, Stanford University, June.

Huitema, C. (1997), "The Required Steps Towards High Quality Internet Services," Unpublished Bellcore Report.

Kazumori, E. (2000), "Rethinking the Role of Government in Economic Development: A Case Study of NSFNET and the Evolution of Internet, 1985-95," Department of Computer Science, Coordination of Agent-based Systems Workshop, Stanford University. Revised: October 1.

Kesan, S. and Shah, R. (2001), "Fool Use Once, Shame on You – Fool Us Twice, Shame on Us: What we Can Learn from the Privatization of the Internet Backbone Network and the Domain Name System, Law and Economics Working Paper Series, No. 00-18, University of Illinois College of Law, Urbana-Champaign, February.

[Available at http://papers.ssrn.com/paper.taf?abstract_id=260834.]

Kleinrock, L. (1961), "Information Flow in Large Communication Nets, RLE Quarterly Progress Reports, MIT, July.

Kleinrock, L. (1964), *Communication Nets: Stochastic Flow and Delay*, New York: McGraw-Hill.

Leib, Volker and Raymund Werle, 1998: Computernetze als Infrastrukturen und Kommunikationsmedien der Wissenschaft. In: *Rundfunk und Fernsehen* 46 (2-3), 254-273.

Leiner, B. M., Cerf, V. G., Clark, D. C., Kahn, R. E., Kleinrock, L, Lynch, D. C. Postel, J. Roberts, L. G., Wolff, S. (2000), "A Brief History of the Internet," The Internet Society, Version 3.31 (Revised 4 August 2000). [Available from: www.isoc.org/internet/history/brief.html.]

Lemley, M. A. and Lessig, L. (2000), "The End of End-to-End: Preserving the Architecture of the Internet in the Broadband Era," Stanford Law School, John M. Olin Program in Law and Economics Working Paper No.207.

Lessig, L. (1999), *Code and other Laws of Cyberspace,* New York: Basic Books.

Licklider, J. C. R. and Clark, W. (1962), "On-Line Man Computer Communication," MIT Working Paper, August.

MacKie-Mason, J. K., and Varian, H. R., (1992/1996) "Some Economics of the Internet," in *Networks, Infrastructure and the New Task for Regulation*, W. Sichel, ed., Ann Arbor: University of Michigan Press. [University of Michigan Working paper 1992, available at: ftp://gopher.econ.lsa.umich.edu/pab/Papers.]

MacKie-Mason, J. K., and Varian, H. R. (1993/1995a), ""Pricing the Internet" in *Public Access to the Internet*, B. Kahin and J. Keller, eds., Cambridge, MA: MIT Press, 269-314. [University of Michigan Working Paper, 1993, available at: ftp://gopher.econ.lsa.umich.edu/pub/Papers.]

MacKie-Mason, J. K., and Varian, H. R. (1994/1995b), "Pricing Congestible Network Resources" (with Hal R. Varian), *IEEE Journal of Selected Areas in Communications*, vol. 13, no. 7 (Sept. 1995): 1141-1149.

MacKie-Mason, J. K., and Varian, H. R. (1997), "Economic FAQs About the Internet," in *Internet Economics*, eds., L. W. McKnight and J. P. Bailey, Cambridge: MIT Press.

McKnight, L.W., and Bailey, J. P., (1997), "An Introduction to Internet Economics," in in *Internet Economics*, eds., L. W. McKnight and J. P. Bailey, Cambridge: MIT Press. ,

Naughton, J. (1998), "To Serve Us All His Days: Tim Berners-Lee, Weaver of the World Wide Web," *The Observer Review*, 19 April 1998, 1.

Norberg, Arthur L. and Judy E. O´Neill, 1996: *Transforming Computer Technology. Information Processing for the Pentagon, 1962-1986*, Baltimore: The Johns Hopkins University Press.

Odylzko, A. (1998), "The Economics of the Internet: Utility, Utilization, Pricing and Quality of Service," AT & T Labs-Research. [Available at: www.research.att.com/`amo/doc/networks.html].

Paxson, V. (1999), "End-to-End Internet Packet Dynamics," *IEEE/ACM Transactions on Networking*, 7(3), June: 277-292.

Rogers, Juan D. (1998),: Internetworking and the Politics of Science: NSFNET in Internet History, *The Information Society*, 14: 213-228.

Salus, Peter H., (1995), *Casting the Net. From ARPANET to INTERNET and beyond*, Reading MA/ Menlo Park CA : Addison-Wesley.

Schmidt, S. K., and Werle, R. (1998), *Coordinating Technology. Studies in the International Standardization of Telecommunications*, Cambridge MA: The MIT Press.

Shapiro, C.. and Varian, H. R. (1999*), Information Rules: A Strategic Guide to the Network Economy*, Boston, MA: Harvard Business School Press.

Zakon, R. H. (2001), "Hobbes' Internet Timeline," Version 5.3,Last Updated, 15 April 2001. [Available at: www.zakon.org/robert/internet/timeline/.]