3-2000

Open-Layered Networks: the Growing Importance of Market Coordination

Petros Kavassalis

Joseph P. Bailey

Thomas Y. Lee
University of Pennsylvania

Recommended Citation

Kavassalis, P., Bailey, J. P., & Lee, T. Y. (2000). Open-Layered Networks: the Growing Importance of Market Coordination. *Decision Support Systems, 28* (1-2), 137-153. http://dx.doi.org/10.1016/S0167-9236(99)00080-9

Open-Layered Networks: the Growing Importance of Market Coordination

Abstract
Based upon the Internet perspective, this paper will attempt to clarify and revise several ideas about the separation between infrastructure facilities and service offerings in digital communications networks. The key notions that we will focus on in this paper are: (i) the bearer service (BS) as a technology-independent interface which exports *blind* network functionality to applications development; and (ii) the organizational consequences associated with the emergence of a sustainable market of BS: a clear movement at the level of industrial structure from traditional *hierarchies* to more *market coordination*.

Keywords
open-layered networks, market coordination, internet

Disciplines
Marketing | Technology and Innovation

Open Layered Networks:
The Growing Importance of Market Coordination

Petros Kavassalis,
Ecole polytechnique, Centre de Recherche en Gestion (CRG), Paris, France
And
Institute for Computer Science (ICS), Crete, Greece
<petros@rpcp.mit.edu>

Joseph P. Bailey,
Robert H. Smith School of Business, University of Maryland, USA
<jbailey@rhsmith.umd.edu>

Thomas Y. Lee,
Massachusetts Institute of Technology, Context Interchange, USA
<tlee@mit.edu>

in Decision and Support Systems 790 (2000)

Abstract (*)

Based upon the Internet perspective, this paper will attempt to clarify and revise several ideas about the separation between infrastructure facilities and service offerings in digital communications networks. The key notions that we will focus on in this paper are: i) the bearer service as a technology-independent interface which exports *blind* network functionality to applications development; ii) the organizational consequences associated with the emergence of a sustainable market of bearer service: a clear movement at the level of industrial structure from traditional *hierarchies* to more *market coordination.*

(*) Note: The authors would like to thank the MIT Internet telephony Consortium for financial support. A previous version of this paper has been published in Bohlin and Levin (eds), 1998, Telecommunications Transformation, Technology, Strategy and Policy, IOS Press, Amsterdam, under the title: Sustaining a Vertically Disintegrated Network through a Bearer Service Market.

1. Introduction

During the few past years, applications like email and the World Wide Web have combined with evolving network protocols to propel the Internet into the heart of a computer and communications convergence. Central to the Internet's immersion into digital convergence has been the effectiveness with which the Internet Protocol (IP) has played the role of "spanning layer." [1].

The IP abstraction enables applications to request network services independent of underlying, physical network technologies. Moreover, new underlying network technologies may either substitute for or co-exist with existing network technologies without significantly affecting the broader system. Based on this abstraction, the National Research Council has articulated the "Open Data Network (ODN)" as an architecture for the networks of the future that generalizes the principle of separating service offerings from infrastructure facilities as demonstrated in the Internet [2]. In the same way that IP serves the Internet, the ODN relies upon a "bearer service" to function as a *technology-independent network layer* that resides above the technology substrate and enables interoperation between diverse, high-level applications and various underlying network infrastructures (figure 1).

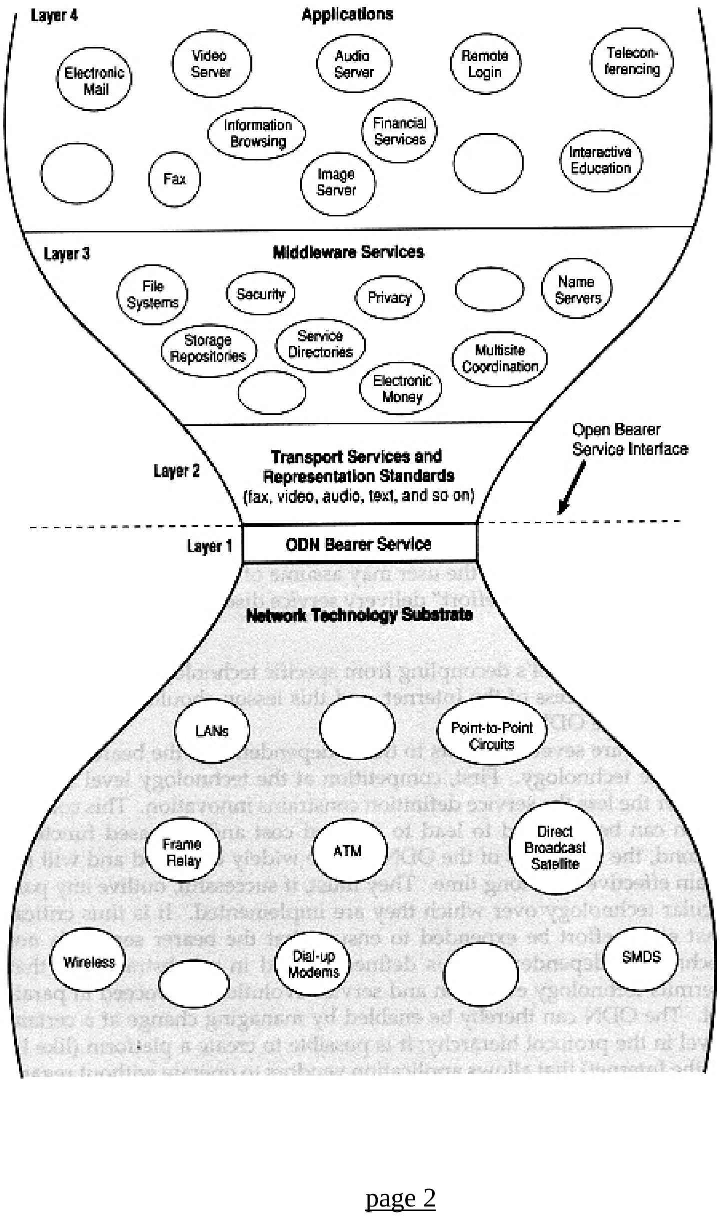


Figure 1. The Bearer Service Concept

The NRC report describes an Open Digital Network (ODN) as a four-level layered architecture: *"i) at the lowest level is an abstract bit-level service, the bearer service, which is realized out of the lines, switches, and other elements of networking technology; ii) above this level is the transport level, with functionality that transforms the basic bearer service into the proper infrastructure for higher-level applications (as is done in today's Internet by the TCP protocol) and with coding formats to support various kinds of traffic (e.g., voice, video, fax); iii) above the transport level is the middleware, with commonly used functions (e.g., file system support, privacy assurance, billing and collection, and network directory services); and iv) at the upper level are the applications with which users interact directly. This layered approach with well-defined boundaries permits fair and open competition among providers of all sorts at each of the layers"[2].*

Certainly, the Internet demonstrates the technical and functional robustness of a technology-independent bearer service abstraction [3]. The bearer service is intended to support requests for service from all applications and to recognize all substrates. However, as both application and infrastructure innovations turn increasingly towards user-oriented models of network architecture, technology and policy considerations related to the generalization of this abstraction should be carefully studied. Such a service blurs the boundaries of telecommunication markets.

For example, the promises of Internet telephony to combine the benefits of the public switched telephone network (PSTN) and the Internet would be possible through a bearer service—even though Internet telephony would weaken market boundaries and challenge the regulatory environment [4]. Regulators have a difficult time categorizing bearer service providers as belonging to any one existing market because such providers can offer services that cut across many existing telecommunications markets. Businesses that develop ubiquitous services or tailor applications for customers may be threatened by market entrants who have competitive products built around a new technology architecture that is able to provide great flexibility in applications design—the bearer service. Finally, customers can benefit from an integrated services environment because their data and voice communications can be transmitted across multiple telecommunications infrastructures. A new era of interoperability [5] is possible through the bearer service.

However, questions for this new market abound. Not only is the business model in question, but the technology is also in flux. In this paper, we span both technical and economic issues to describe how the bearer service may be provided through markets and not hierarchies. We do this by shedding some light on the many questions arising from this new bearer service market. First, what does a technology independent bearer service look like? In what ways is this technology different from older telecommunications architectures? How these differences are interpreted in *organizational terms?* Second, is a segregated network that separates infrastructure facilities from service offerings by a bearer service an economically viable market? And if so, should we expect *market coordination* becoming more important than *organizational coordination* in the value-process of modern communications industries?

The remainder of the paper is organized as follows.

Section 2 employs a comparative analysis to define concepts of the bearer service and the *layered network architecture* concepts. Considering IP as a *bifurcation point* in the evolution of network design, the bearer services of traditional communications infrastructures and the Internet (both the current *best-effort* and the future *Integrated Services* Internet) are surveyed to elicit design characteristics and functional differences between a technology-dependent and independent bearer service. We suggest that in a network design with a technology-independent bearer service, the communications network supports a *flexible organizational* model capable of dealing with new and unanticipated applications.

Section 3 associates the technology-independent bearer service with the Internet organizational model: a flexible system of regional or more extended backbones and access links (or access networks) to these backbones, managed by the Internet Service Providers (ISPs). Specifically, we address the question of whether the ISP model, which is characterized as a model where network operators exhibiting varying degrees of vertical integration compete in an open market, can sustain itself, or if one monolithic, integrated firm will emerge from mega-mergers? To answer this question, we begin by considering the work of Gong and Srinagesh [6, 7], who argue that a stable and sustainable equilibrium for healthy bearer service market growth might not be possible because the firms that own the substrate network layer will monopolize the bearer service market. Our analysis closely follows the definition for a bearer service formulated in section 2. Arguing that the bearer service is not a commodity product, we identify differentiable service attributes upon which an independent bearer service market could form. Through differentiation, the bearer service market can avoid a Bertrand equilibrium (i.e. pure price competition). Furthermore, we challenge assumptions about perceived trends towards vertical integration, by noting the relative independence of bearer service assets and underlying infrastructure facilities.

2. From the railroad gauge to information bitways: the evolution of the bearer service functionality

Modern communications networks, such as the Internet, are technologically heterogeneous and decentralized regarding the distribution of the network intelligence. The only common denominator bridging over an increasing structural variety is the *bearer service interface*. Given a pre-specified set of applications and a physical network which may include more than one substrate technology (e.g. a Ethernet-based LAN connected to a Frame relay metropolitan network), the bearer service (BS) constitutes those common[1] functions which are implemented throughout the network rather than in

---

[1] If every application uses the function, then it is certainly a function in common and unambiguously a component of the Bearer Service (BS). If only one application uses the function, then perhaps it is more appropriately considered part of the application, but it does not span the many applications so it cannot be part of the BS. If two or more applications utilize the function but not all applications in the set use the function, then we need to question whether the function belongs in the BS. Recall also that a separate metric for distinguishing BS functionality is whether that function can be

the network end nodes and are necessary for pairing each application's communication requirements with the performance characteristics of all components of the heterogeneous network.

More abstractly, computer and communications technologies may be separated into three layers. The physical infrastructure (e.g. wires, switches, etc.) resides at the lowest layer. At the top lies the set of applications and service offerings supported by the underlying infrastructure. A *spanning layer*[2] bridges the two [1]. An application requests network services through the spanning layer to the substrate technologies. In the early public switched telephone network (PSTN), telephony was tightly coupled to a specific infrastructure so the spanning layer supported only a single application and one technology substrate (typically, the spanning layer was located *in the wires)*. The development of newer applications for computers and communications as well as advances in substrate technologies, prompted a refinement of the spanning concept. In the presence of a diverse suite of applications and a heterogeneous network, the *bearer service (BS)* constitutes a spanning layer which *escapes from the wires* and thus supports all applications over the entire network as long as the Bearer Service is able to pair a service request with an underlying substrate technology end-to-end.

This section uses Piore's model of organizational flexibility and production system transformation as a methodological framework for tracing the evolution of the spanning layer towards a technology-independent bearer service. Advances in shipping and the transport of physical goods are used as a metaphor for the transformation of yesterday's PSTN into tomorrow's ODN.

2.1. Production technologies and flexible specialization

Piore [8, 10] describes the on-going transformation of industrial production systems towards greater variety and flexibility as a four-stage evolution[3]. The products in such systems are comprised of both independent and interdependent design features; changes in design features mark the different evolutionary stages. Independent design features "can be varied in isolation without complementary changes in other features of the design" while interdependent features "require a number of complementary adjustments" [8].

The initial stage, *mass production* is characterized by a production system tailored to a single product. There is no room for variation. In *Mass production with cosmetic variation*, product design may slightly vary existing or may introduce new independent design features. *"The notion of cosmetic variation seems to imply a sharp dichotomy between design changes which are easy to make and those that are not" [8].*

---

implemented in an endnode. BS functions include only those functions that cannot be implemented in an endnode.

[2] As suggested by Clark, "a spanning layer is characterized by three key parameters that characterize the sort of interoperation it supports: i) the span of infrastructure options over which the interoperation occurs, ii) the range of higher level applications the spanning layer supports, iii) the degree of symmetry in the services of the spanning layer and its interfaces, and in the higher layers up to the application definitions". [1]

[3] This work draws on *The Second Industrial Divide* [9]

Flexible mass production extends cosmetic variation by introducing the potential for change in interdependent design features. Flexible mass production explicitly identifies, a priori, both the set of product design features which is subject to change and the set of values which each design feature may take. Therefore, the flexible mass production system represents a finite number of products, which vary in more than simple cosmetics.

Diametrically opposite mass production is *flexible specialization* where variation is virtually infinite. However, *closed flexible specialization* includes those systems where the set of design features that varies is defined a priori, but the domain over which each varying feature ranges is unknown. By contrast, *open flexible specialization* where both the set of variable design features and the domain over which each variable ranges is potentially infinite.

A different cognitive model applies in each stage, with flexible specialization involving a balance between *"a deepening of understanding within a given cognitive frame and the pull to reintegrate back (in the production process) to a different frame in order to produce a sellable commodity" [11].* Similarly, a technology-independent bearer service offers more than a pre-designed set of services. Rather, it supports an "*application-blind interface*" that enables the introduction of new applications independent of the initial strategies and service offerings of Telecommunications Operators (TOs). This functionality can be easily reintegrated to the application vendors and users' cognitive frames, thus allowing them to introduce new applications and operate independent of the strategies and the service offerings of the Telecommunications Operators (TOs). To illustrate the continuum that spans well-defined, mass production systems and flexible, application-blind interfaces, we will discuss two contrasting metaphors, the *gauge* and the *container.*

2.2. The gauge metaphor

The gauge of a railroad is defined as the distance between rails or between the flanges of the wheel sets on a railroad car. The gauge determines the tracks upon which a given railroad car may travel. By extension, the gauge therefore also determines which railroad companies may exchange rolling stock and the transparency with which a customer may transport freight across boundaries between different railroad companies. Accordingly, diversity in gauge standardization implies transaction costs and other inefficiencies as customers and freight traverse rails of different gauges. Thus, for reducing technical complexity and transaction costs and for internalizing mutual network externalities, railroads have been progressively converging towards a gauge standard [12, 13].

The emergence of the spanning layer concept may be derived directly from the convergence towards a rail gauge standard. Railroad tracks comprise the physical network. Differences in rail cars represent distinct applications from which a customer might choose. Gauge dimensionality is therefore a layer that resides between the physical track and the applications. Gauge standardization expanded the services that a particular rail system could offer by extending the reach of the rail network and expanding the scope of traffic (the kinds of cars) that could be carried. Standardization reflected a shift towards some flexibility where rail car design could marginally vary (as long as it conformed to the gauge standard) and traffic could move across network boundaries.

2.3. The spanning layer in traditional communications networks

Communications networks have traditionally been vertically integrated. Whether for telephony, radio, broadcast television, or community antenna television (cable television), infrastructures have long been closely coupled to service provision. As Tennenhouse et al. notes [14], *"telephone and cable services are each carried over their own wired systems. Although radio and television share the airwaves, for practical purposes, they are discrete distribution vectors, since separate portions of the spectrum have been allocated to each type of service"*. With one wired network for any given application, each service resembles a mass production system. *Spanning* is not obvious because it is impossible to distinguish the layers of a monolithic system.

Even within a single service offering, however, the utility of a spanning layer is apparent. Telecommunications heterogeneity is subject to the same economic forces promulgating rail gauge convergence. Monopoly structures and the political environment contributed to both the rapid acceptance of the 4 kHz circuit as a worldwide standard for voice communications[4]. Subsequently, that standard emerged as a spanning layer providing transparent support for multiple applications (for instance fax and data transmission via modems) over technologically evolving analog and digital telephony[5].

The advent of digital computer and communications technologies, in addition to introducing the prospect of new applications and services, began to extend the scope of the spanning layer concept. For digital telephony, the 4MHz channel gave way to a transfer rate of 64 Kbps using pulse code modulation (PCM). While corresponding to a single voice conversation, the data rate of 64 Kbps also introduced a new category of *value-added* applications and services (e.g. credit-card calls, file transfer, audio-conferencing, etc.).

For value-added services, however, a raw 64 Kbps bit pipe proved inadequate. Specifications for data format, routing, signaling, and many other parameters of transmission are required to enable high-speed data services. Such services include packet switching, Frame Relay, intelligent network applications, and Integrated Services Digital Network (ISDN). Consequently, value-added services led to a new network model that extended the 64 Kbps spanning layer with delay, loss, and other data characteristics. Rather than cosmetic, these additions drove the system towards flexible mass production where the network supported a discrete set of delay values, etc. to support a limited number of new, pre-specified services.

The additional switching and control functions for these new services are *logically* implemented on top of the elementary 64 Kbps spanning layer yet *physically* implemented elsewhere than the Central Office Equipment. As examples of the distinction between the logical and the physical, consider X25 and ISDN. The

[4] As in the case of the railroads, however, those forces may sometimes be unable to achieve complete standardization. In broadcast television, *spatial lock-in* contributed to the emergence of three different standards (NTSC, PAL, and SECAM).

[5] The development of applications other than those selected by the network operator, though technically feasible from the beginning, became however effective only following a change in political climate as signaled by the FCC's *Carterfone* decision [15].

transmission of packets between nodes attached to the same transmission link is assured by a specific *connection-oriented* protocol, the X25[6]; the *data link* layer of the X25 converts an *unreliable* for packet transmission bit pipe to a packet-link, and, in collaboration with the *network* layer, supervises naming, addressing, routing and congestion control[7] [16, 17]. In a similar way, the ISDN (2B+D) provides a customer with two B channels of 64Kbps[8] and one X25 full duplex D channel at 16Kbps for "out-of-band" signaling; furthermore, under the *Intelligent Network* model, ISDN separates the flow of control information from user information[9] [18, 19].

Despite the new functionality and extensions to functionality supported by the spanning layer, the spanning capability was essentially limited. While a single infrastructure may have supported more than one application, service offerings continued to be tied to the infrastructure in a largely many-to-one correspondence. Infrastructures were not application blind. For example, changes in the X.25 protocol would necessitate change both within the network and at the end nodes. Overall, the public network resembled the shape of a "patchwork" of vertically integrated, flexible mass production systems where each network supported a different set of applications.

Within each of these production systems, one can not distinguish between the conception of a service and its delivery, because of the spanning layer is *being embedded* in the substrate technology. Variation exists but only the variation that has been ex-ante conceived and designed by the engineers of the infrastructure-facilities owners…

2.4. The new metaphor: the container

Early transportation systems also provide examples of vertically integrated systems. For the purpose of transporting goods over land, dedicated technologies including railroad tracks, cars, and rail yards were created. Likewise, air and water transportation services warranted similarly single-purpose infrastructures for ships and airplanes.

Over time, however, it became common for freight transporters to link different media, such as water, air, or land. Intermodal freight transportation refers to the linear combination of two or more transportation services. For example, moving a product from a warehouse in Hong Kong to a retailer in the American city of Chicago might entail trucking the product to the harbor, shipping the product to a port on the West Coast of the United States, and then conveying it by rail to Chicago.

Contrasting vertically integrated architectures, where a dedicated infrastructure is used to realize separate applications, is the *horizontal integration* model [22], where diverse applications are developed independent of any underlying architecture. Instead,

---

[6] Other technologies for *connection oriented* packet switching include SNA of IBM and DECNET of Digital Equipment.

[7] We refer to OSI terminology

[8] These channels can be indifferently used for voice and data transmission

[9] Even though theoretically possible and very appealing to separate user and flow information, this separation *(unbundling* in the telecommunications jargon) is actually difficult to implement, for reasons relating to the particular strategic interests of the Telecommunications Operators [20, 21].

developers build to a common *abstraction*[10]. The common abstraction, a spanning layer now located *on top* of the transportation (or communication) modes, projects a virtual infrastructure which itself comprises one or more interoperable physical networks. The spanning layer pairs applications to infrastructure facilities because applications request network service[11] from the spanning layer not the underlying infrastructure, and the spanning layer translates a request for network service to the infrastructure protocols.

For the transportation industry, horizontal integration, which emerged only in the 1950's, was made possible by the development of and agreement upon a standard freight container. Muller noted that *"when it was publicly demonstrated in 1956 that standard containers could move successfully on a land-sea intermodal journey, a commercial revolution was started… It was the container's unique role as common denominator among modes that was revolutionary* (we underline)" [23].

2.5. The Internet Protocol: a robust technology-independent spanning layer

In the communications industry, diversity and, consequently, requirements for a common denominator among different technological media, also become increasingly important. Local and Metropolitan Networks providing inexpensive and fast interconnections for personal computers and workstations proliferated inside and outside firms and among research institutions and universities. But, successive technological waves have dispersed the computer networking environment in different architectures (LANs: Ethernet, Token Ring, Apple Talk, MANs: Frame Relay, SMDS etc.). Interconnection between these architectures, as well as connections between these private networks and the public telecommunications infrastructure, appeared as the natural priority. In the 80s, three technologies were compete to capture the interconnection market (SNA, B-ISBN, ARPA's Internet protocols), with the *Internet protocol suite* becoming more recently the solution most often adopted [24].

The Internet approach to interconnection involves the separation of infrastructure bitways from applications by defining an interface (Internet Protocol) to the basic infrastructure facilities and then exporting that interface for application development. Formally, IP provides a uniform method of addressing (which is independent of physical hardware addressing) and a variable size datagram *(i.e. a standard data container)*[12]. From this interface any number of diverse applications (e.g. the World Wide Web as an example of new, unanticipated applications) may be constructed [25]. Likewise, principles of datagram encapsulation enables any number of substrate technologies to transmit IP (e.g. Ethernet, token ring, X25, ISDN, Frame Relay, ATM, SONET, WDM) [26].

---

[10] Messerschmitt [22] defines *abstraction* as "the conscious hiding of unnecessary implementation of functional details while making visible behavioral properties that are essential and important to other modules. Abstraction helps to ensure the independence of the architectural modules, as one module cannot be dependent on the hidden properties of another module".

[11] The term "network service" indicates here the actions and responses (i.e. functionality) provided by underlying infrastructures to application requirements. Not confuse this *network service* with that we usually call *services (or service offerings),* i.e. the customer services provided by a network infrastructure to final users.

[12] To use Rose's terms [26], *IP has no wires associated with it.*

IP takes an additional step towards supporting flexibility by disassociating *transport* from *network* functions (respectively performed by TCP and IP) [25]. As a result, IP incorporates only that functionality which must be implemented within the network as opposed to at the periphery or end-user nodes. In this way, IP minimizes the possibility that changes in applications (including transport protocols) or infrastructure components will require changes in all of the routers, hubs, and gateways throughout the network substrate.

By separating spanning layer development from both infrastructure facilities (and applications), IP characterizes a different technological trajectory[13] that stems more immediately from the container metaphor. Essential characteristics of this trajectory are that applications may work over multiple substrates (e.g. network technologies) and that these substrates do not pre-specify the development of new applications. The separation both above and below, move IP further along Piore's framework towards flexible specialization: IP offers flexibility[14] and features variability both above the spanning layer (e.g. applications) and below it (e.g. technological substrates).

However, the current IP configuration provides only a *best-effort* delivery service. Consequently, it can not deal with applications requiring service guarantees, such as real-time applications (telephony, video etc.). To respond to this objective, new service models are being defined in the Internet together with protocols to reserve capacity according to applications' requirements or by giving precedence to certain categories of traffic or users, or both [29]. Essentially, the goal is to bring out an architecture for Integrated Services Packet Networks–ISPN [30].

New concepts for spanning, therefore, propose to complement IP datagrams with functions to provide not only best-effort delivery but also variable Quality of Service (QoS), so packets can get to their destination quickly, consistently and reliably. With the major part of the research effort devoted now, to create a new "building block" able to specify QoS parameters (as bandwidth, delay and loss characteristics), the search for a new Internet architecture seems to cognitively approach the ODN model. Both approaches impose a narrow point in the protocol suite, *"isolating the application builder from the range of underlying network facilities, and the technology builder from the range of applications"[3].*

2.6. From spanning layer to bearer service

This section began by considering the traditional telecommunication infrastructure and its successive evolutions in light of Piore's perception of the transformation of the industrial system. The organizational archetype is the *Bell System* design, a ubiquitous *mass production* system for telephony, where application and substrate are not only tightly coupled but also in fact developed in concert. In time, this model has produced a slight

[13] We us the term technological trajectory as defined by Dosi [27]

[14] Borrowing from Bar [28], "... it is the flexibility of its network resources will determine a company's ability to experiment with telecommunications technologies, learn from the experimentation, and repeatedly reorganize itself to capture cumulative benefits. True network flexibility is not only the ability to support a range of applications over a given network configuration but also, and perhaps more importantly, the simultaneous ability to re-configure a network in order to provide various applications mixes and to design new applications that take advantage of new configuration possibilities".

differentiation, by offering a number of additional services (fax and data transmission via modems) over the initial spanning layer (i.e. a 4 kHz point-to-point channel, the telecommunication equivalent of the *railroad gauge).* With the introduction of digital technologies, we observed a more fundamental transformation in network production towards increasing variety at the level of service offerings. However, the application and service design are still designed into the system ex ante. Therefore, flexibility is *formally* developed within a *"closed set with a finite number of elements" (flexible mass production).*

The section then explored newer network models. Interconnection of heterogeneous networks (as exemplified by the Internet) and, now, horizontal integration, mark the continuing shift from *cosmetic variations* in mass production to more *flexible mass production* and, more recently, to *flexible specialization.* In the newer models, applications may be able to go over multiple network technologies (substrates) without the need to design applications and services in advance, as a function of particular substrate constraints.

By fitting the evolution of network architectures to Piore's framework (figure 2), we were able to extract functions of the *spanning layer* and predict how the *bearer service* may evolve. A trivial spanning layer, embedded in the substrate technology, seems to characterize early railroads and telecommunications mass-production networks. Increasing service variety, due to digital transformation of the network, required to develop extensions to the functionality supported by the spanning layer. However, the spanning layer itself remained tied to the infrastructure *wires.* It escapes from the wires with the Internet, which proposes a network model that might be described by the *container metaphor:* a specific protocol (IP) plays the role of a common denominator among different substrate technologies. Because IP is a technology-independent spanning layer which enables independent variation of applications above it as well as substrate technologies below it, it is at a layer above the bitways which enables more design flexibility and innovation.

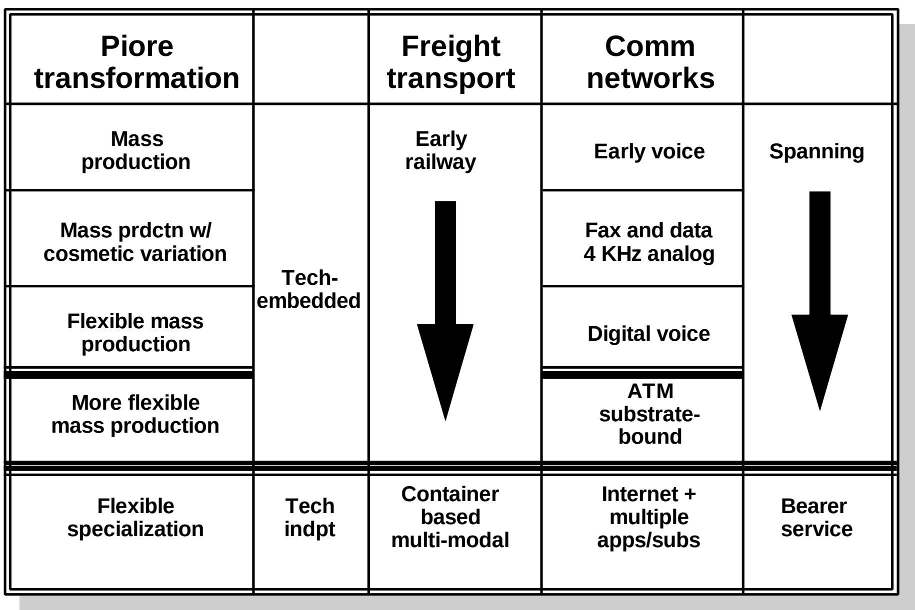


Figure 2. Trajectories of technical change

In the near future, integrating flow control mechanisms (assuring Quality of Service) with IP should extend Internet's application variability to include services with real-time constraints. Again, rather than implementing this functionality within the bitways, an enhanced IP embeds the additional functionality into the spanning layer. In this way, a new "building block" appears for supporting a network composed from Integrated Services Packet Networks (ISPN). These characteristics of the new Internet "building block" (the span of underlying infrastructure facilities and the range of applications that can be served) in combination with the existence of *symmetrical* properties (the type of delivered service is enforced by the end points) make it a close manifestation of the ODN *bearer service* concept for two reasons. In the ODN network model, the discussion on the bearer service is in abstract terms and there is no provision for implementation. But the position of the "spanning layer" is exactly the same as the bearer service, it is close to the network technology substrate. Throughout, the whole system may scale over number of networks and different substrates. And second, the ODN bearer service layer is clearly defined as providing quantified Quality-of-Service measures for bandwidth, delay and loss characteristics. Simultaneously, this layer must also provide mechanisms for accounting at a sequence of packets level and incorporate feedback loops to utilize accounting statistics in capacities such as evaluating QoS commitments and traffic management.

This approach contrasts with an "all ATM (Asynchronous Transfer Mode) network" strategy for horizontal integration. Though ATM uses a common container (a fixed size cell), in at least one respect, ATM appears to evolve directly from the *railway gauge* metaphor: while ATM is application independent, it is still firmly linked to the substrate technologies. As a result, evolution paths should be narrowed to ATM upgrades. From this point of view, ATM tends more towards a very *flexible mass production system* (or a *flexible specialization with large and discontinuous change* system[15]) by pre-specifying the basic commodity products for meeting customer needs or constructing higher level applications (ATM was designed as an applications independent platform with three general applications at its center: telephony, high speed data communication and video delivery). In the history of technology, ATM will be perceived as an excellent infrastructure technology for supporting bearer service's functionality.

To conclude: A spanning layer lies between a set of network substrates and higher level applications. *One special case of a spanning layer is the bearer service that is both substrate independent and application blind.* Only those functions that cannot be implemented at the periphery of the network substrate are included in the bearer service. Because not all substrate technologies can support all application requirements, application requests must be matched to particular substrate characteristics. To preserve transparency, the bearer service conducts the pairing rather than having applications explicitly request particular underlying technologies. This pairing is related to *a flexible specialization* model for network organization where applications and services may

---

[15] Piore [8] defines *flexible specialization with large and discontinuous change* as follows: "In any system, some moves will be incremental and some will be large and discontinuous... An open (flexible system) for which a change is large and discontinuous is in some respect similar to a system of flexible mass production".

instantaneously *emerge,* without the need to have designed them in advance, at the same time when infrastructure technologies have been developing.

3. The organizational consequences of an independent market for bearer service: more market coordination

The evolution of communications spanning layers from homogenous substrate/application pairings to the bearer service may change the economics of this industry. While section 2 concentrated on how spanning layers have evolved to incorporate heterogeneous applications and substrates, this section will explore the economics of this evolution and describe the eventual structure of the communications industry after the Internet. We will describe why a bearer service market may be sustainable and if so, enable industry agents to flexibly re-organize firm-specific assets and easier differentiate their strategies and offerings. Essentially, we will raise questions about possible *governance structures* in networks with layered architectures where layers correspond to independent markets.

3.1. Institutional forms associated with an independent bearer service layer

The argument we developed earlier built around a distinction between two different trajectories of technical change.

The first derives from traditional telecommunications *mass production* organizational models and involves a pattern of innovation consisting of anticipating user needs in communications services and designing the corresponding underlying network architecture.

In contrast, the Internet and newer approaches for horizontal integration of an infrastructure becoming increasingly heterogeneous and separate bitways from applications (and service offerings). Technically, this is possible by defining an interface (IP) to the basic infrastructure facilities and then building on top of that interface the general functionality for applications. We argued that this particular approach to technological change leads to a new *technological paradigm.* Its essential difference from previous organizational models is that applications may work over multiple substrates (i.e. network technologies) and that these substrates do not determine the development of new applications. It is increasingly clear that this new trajectory brought up considerable flexibility in the design of applications the customers use (as exemplified by the extra-ordinary development of the World Wide Web). Because it “blindly” uses network resources (i.e. flexibly shifting network resources across different uses), system production (applications and service offerings) takes the characteristics of an open-ended product line able to generate (possibly infinite).

As usual, technologies and economic institutions are evolving together [31]. Indeed, the emergence of a new intermediary stage offering *IP connectivity* by using input provided by various infrastructure facilities (leased lines with fixed capacity as well as Frame Relay, ATM or SONET connections and now IP over WDM) has several institutional consequences. First, a market system has emerged to coordinate the relationships between sellers and buyers (e.g. intermediary providers or final customers). In this market, owners of infrastructure facilities offer raw or “cloud” network capacity to Internet Service Providers (ISPs) which, in turn, sell IP-packet transport either directly to the customers or to other ISPs, Internet Access Providers or Service Integrators. In many cases, facilities’ owners and ISPs belong to the same vertically integrated company. In other cases however, independent ISPs may rent facilities to large integrated companies (MCI, Sprint etc.) or to specialized bandwidth providers (as Qwest, Level 3 etc.). This *vertical* market exists together with other *horizontal* markets for *IP interconnection* between interacting ISPs linked through *settlements* [32,33]. In that way, the Internet seems to favor the emergence of wholesale markets for IP transport, on top of the traditional bandwidth markets for raw capacity (fixed leased lines) or switched transport.

Then, the question easily arises: should we expect the stabilization of these trends to a global market for IP connectivity? It is clear that the survival of the *flexible specialization model* requires the development of these markets that therefore can support the co-existence of multiple business strategies – a mix of vertically integrated and independent ISPs based on external provisions of network capacity. Or, will the current dynamism of the Internet industry finally be absorbed by the resurgence of the *flexible mass production trajectory* through the revival of an industry structure dominated by few large vertically integrated firms? However, before replying to these questions, we must address the question, "Are these emerging markets for IP connectivity and, furthermore, for bearer service functionality, sustainable or not?"

3.2. Unbundled bearer service?

A market for a technology-independent bearer service may not be sustainable, according to Gong and Srinagesh [6, 7]. Their argument consists of a series of propositions that we summarize here:

i) In networks with layered architectures, competition at the bottom layer (infrastructure facilities) of the network hierarchy is unsustainable. Markets for raw transport under conditions of excess capacity and oligopolistic competition for a homogeneous good, turn easily to “destructive competition” arising from Bertrand equilibrium[16] (prices decrease to marginal costs following the example of the leased line market). Consequently, facilities-based carriers competing for raw transport essentially on price, may not cover their sunk and fixed costs and fail to afford competition with non-facilities based resellers and ISPs.

ii) One way to avoid Bertrand competition is through bundling of bottom-layer transport with higher services, closer to final customer (vertical integration). With bundling, carriers are able to differentiate their services, segment the market and price accordingly.

[16] Gong and Srinagesh describe raw transport as a service that is like a commodity. A commodity is a product that is so standardized and sold by enough firms that no firm can set the price — all firms are price takers. The Bertrand competition model in economics holds for such a market. Assuming that all firms have a different marginal cost and they will exit the market if the price is set below or equal to their marginal cost, the only sustainable market price is equal to the marginal cost of the second-lowest firm.

iii) As facilities-based companies integrate with others at higher layers, variable costs rise significantly, resulting in a U-shaped average cost curve with a minimum efficient size that is small in relation to the size of the market. As a result, more firms are supportable in equilibrium under vertical integration conditions.

iv) Given the above hypothesis, *"policies promoting competition in the provision of un-bundled bearer service among owners of physical networks may ultimately fall"*.

Gong and Srinagesh extend their analysis to other forms of bundling and horizontal mergers to demonstrate that integration may be a natural outgrowth of competition in the convergence spurred by technological advancement (layered digital architectures) and deregulation. Albeit increasing diversity in service offerings, *economies of scale* and *sub-additivity* seem still powerful and responsible for generating growth, as in the old trajectory of *mass production.*

However, counter examples from the Internet come easily to mind. A quick examination of *The List* (http://www.thelist.com), a listing of Internet Service Providers, offers many example of non-integrated ISPs, with good local or more extended implementation or with specialization in business users. Kavassalis and Lehr [34] recognize the existence of disaggregation trends in the emergence of new Internet providers as for example Qwest and Level 3, with strong position in the bandwidth markets, or Savvis and InterNAP offering new business models for Internet provision. As they note: *"Qwest is the best example of a specialized provider (with a very large and fast fiber network across the US), offering abundant bandwidth in the wholesale (carriers and ISPs) and retail markets (i.e. directly to business customers or individual consumers – IP telephony). Savvis and InterNAP are the more prominent examples of a new kind of ISP, local aggregators. They aggregate traffic from other ISPs or web-based providers (as Point Cast or CDNow) and then, forward it quickly to the big national backbones (MCI, Sprint, UUNET, ANS etc.) through private NAPs (a sort of "aggregate hubs", concept similar to airline hubs). Private NAPs-based companies obtain much better performance and reliability because, i) they avoid highly congested public NAPs and MAEs (where ISPs interconnect each other) and, ii) do not rely on one only backbone".*

By critically discussing the arguments of Gong and Srinagesh, the remainder of this section will focus on the reasons justifying the sustainability of a market for IP connectivity and bearer service. In the beginning, we will explain the characteristics of the "bearer service" *product* (very different from a commodity product) and explore opportunities for a sustainable bearer service market under two different perspectives, i) *product differentiation* and ii) allocation of *rights of control.* Next, we will put forth our perspective on the issue of *markets versus hierarchies* regarding governance structure in the Internet industry.

3.3. Sustainability of the market

As we have already mentioned, in considering the economics of the bearer service layer, Gong and Srinagesh (op.cit) make the analogy with raw transport markets. The commodity nature of these markets, they argue, make prices go constantly down, so bandwidth providers try to build long-term contracts with the customers and differentiate through individual customer-support and maintenance. This is certainly true but it is not likely to provide a good example for discussing bearer service's economics. Mainly, for two reasons. The first relates to the proper economic of the private lines market and the second to the different nature of the bearer service product that does not look like a commodity.

1. Prices in the private lines market are strongly dependent on the availability of fiber-laying capacity that may fluctuate as a function of the investment cycles. After ATT's breakup for example, the telecommunications industry massively invested to build new fiber backbones. The resulting bandwidth excess has driven down the prices for private leased lines until the beginning of 90s. In fact, this trend will be reversed by the emergence of the Internet and the new demand for bandwidth created by the World Wide Web and the proliferation of the Intranets. Current massive investments in Wave Division Multiplexing (WDM) technologies and the emergence of the new carriers specialized in bandwidth provision (as Qwest and Level 3) may again drive the prices down but the costs also are decreasing[17]. So, no clear conclusion about "destructive prices" may be drawn. Besides that, long term contracts and non-linear pricing are simply two forms of quantity-dependent price differentiation and they may not correspond to any "vertical restraints" imposed from bandwidth providers to their customers, as Gong and Srinagesh implicitly consider[18].

2. Defined as a network substrate-independent interface with Quality of Service (QoS) characteristics (section 2), the bearer service *as a product* looks like *sheer bandwidth with QoS attributes,* available over wholly owned networks as a standard (platform-independent) good with quality options (QoS^1, QoS^2, ... QoS^n). From this definition it should be inferred that bearer service is different from a commodity product and, consequently, the bearer service market would be different from raw transport markets.

A commodity product has the characteristic that the demand elasticity of substitution is infinite. This particular characteristic of a commodity results in the service providers (or vendors) to set their price at the same level[19]. The leased line market may not have an infinite demand elasticity of substitution, but it is much larger than the elasticity for substitution of a bearer service product. The reason comes from the ability of the leased line vendor to fulfill the service request it receives from the customer versus the bearer service provider to do the same. In the leased line case, customers specify the bandwidth of a connection often days or weeks in advance of using that bandwidth. There may be some specification about the reliability of that link, but, in general, the leased line is always there for the customer. In a bearer service market, the customer demands service in a dynamic manner dependent upon the application's service requirements—perhaps only a few milliseconds before the service is delivered. Therefore, the bearer service

---

[17] See *Business Communications Review,* August 1998, pp. 12-14 (J. Puttre, The Oncoming Glut of Bandwidth). The paper reports that T1 prices increased 13 per cent in 1997.

[18] Gong and Srinagesh consider long term relationships between sellers and buyers as an example of vertical integration by providers into the services layer (i.e. out of a bearer service layer), in search of differentiation. However, M. Katz [35] defines as "vertical restraints" only those provisions of an intermediate good which make the buyers' payment to a given seller depend on variables other than the quantities of the input purchased from that seller, statement which is adopted in this paper.

[19] Assuming slight product differentiation, the market would be shared by a number of companies and an equilibrium price would be established — given different product characteristics and availability.

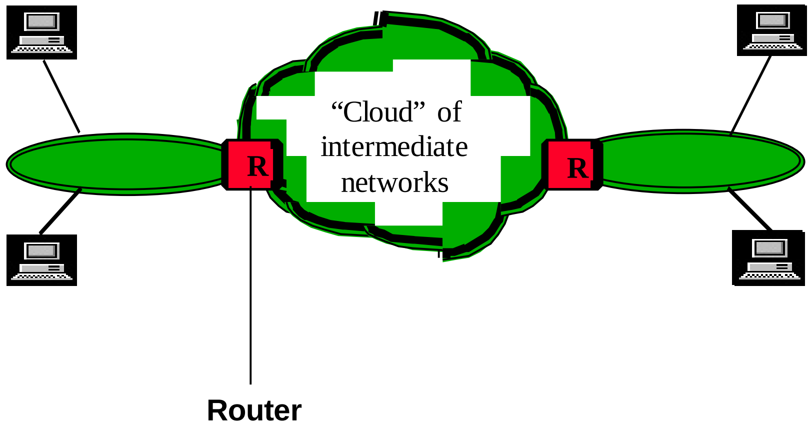
provider must choose technology substrates to service application requests based upon the expected demand. This not only results in a usage sensitive pricing policy to limit the use, but it also encourages the service provider to make the exact provision in network capacity to minimize costs (taking advantage of statistical sharing to service their users requests). The result is a probability of blocked service request that does not exist in the leased line market. Customers then see a noticeable difference between bearer service providers resulting in a price versus performance decision on their part. Because one bearer service provider underprovisions its network less than its competitor, it will also increase its performance by refusing fewer service requests, thereby commanding a greater price for its service. The result is a demand elasticity of substitution that is less than the leased line market indicative of a non-commodity product.

Another commodity characteristic that leased lines have and the bearer service product does not, is transport between two points as opposed to between a point and a shared "cloud" (figure 3)[20]. The leased line customer specifies the points they wish to connect, not only their bandwidth requirement. It is a point-to-point connection. In the Internet, however, the service between two points traverses through the cloud and the full point-to-point performance is dependent on the sender's access link *and* the recipients. Likewise, the bearer service does not specify two communication points a priori, but it only gives access to the shared cloud. While the bearer service cloud does more than the Internet cloud (e.g., provides a guaranteed quality of service), the ultimate functionality depends upon the access link of bearer service provider connecting the sender to the cloud and the quality of the access link connecting the recipient of the cloud. If the same bearer service provider connects both the sender and the recipient, there may be no reason to traverse the cloud and the provider can handle the transmission internally. The result is a provider that is able to offer better service because it has a large network of customers. The economics of *increasing returns* [36] apply here giving customers greater value if there are more customers connected to their bearer service provider's network.

Overall, the leased line product does look more like a commodity than a bearer service because the offering is relatively homogenous and static. The service demands of customers in the leased line market do not change dramatically from day-to-day because the service has been established a priori for a fixed time period (usually months or days). The bearer service provider must service requests very dynamically and it must service requests for an unknown or short period of time (perhaps seconds or minutes). Furthermore, the leased line product that one provider can offer versus another provider is not very different. In the bearer service market, there are distinguishing characteristics about the provision of bearer service that can differentiate the providers, thus making competition less susceptible to "destructive prices". In addition, bearer service provision requires complex investments in different families of assets ranging from network hardware and management tools to link capacity.

---

[20] This difference also applies between the leased lines market and the current *best-effort* IP connectivity market.



The "cloud" metaphor is used to describe the statistically shared wide-area transport of the Internet. In this paper, we argue that the "cloud" metaphor can also be used to describe the bearer service network topology.

Figure 3. A topology for bearer service provision

3.3.1. Bearer service differentiation

There are many pragmatic reasons why the bearer service should have a lower demand elasticity of substitution relative to a commodity, which result from the provision of bearer service. Consumers may take into consideration the firm they are transacting with and not just the product they are buying when they value an exchange. For example, a firm that offers guarantees or has a trustworthy reputation may actually be able to sustain a higher price for a good that is homogeneous. Because consumers value the good as well as their relationship with their transacting party, we can envision many dimensions that bearer service providers can differentiate their service. In fact, bearer service providers can: i) choose different substrate technologies, ii) design a different network topology than their competitor and, iii) design a different pricing policy for their service.

The choice of substrate technology is the first way that bearer service providers can differentiate themselves from a competitor. As discussed earlier, while the bearer service can support all applications and all substrates, it is not possible for all application/substrate pairings to work together. But, by choosing a particular substrate technology, the application space can be constrained. For example, if the bearer service provider chooses Ethernet as its substrate technology, it cannot offer a guaranteed quality of service to the customer while their competitor with an ATM substrate can offer this service. The subset of applications offered by the Ethernet-substrate provider is smaller than the ATM substrate provider is, but their cost structure is also different. By choosing Ethernet, their costs are lower and, therefore, they can pass on some savings to the customer. In summary, the application space may be constrained by choosing a particular substrate technology.

The choice of network topology is another degree of differentiation between the bearer service providers. As outlined above, connection to the *cloud* is very different than connection between two points. Therefore, the way a bearer service provider connects to the cloud and how dependent the provider is on the cloud for service are differentiating factors. The better the connection to the bearer service cloud and the less dependent the provider is on the cloud, the lower the probability of a denied service request. To use a metaphor from transportation, the bearer service provider can provide transport to a "hub" or exchange point where you can hop on another link. If this exchange point is in a remote location, or experiences heavy congestion, then the service is providing you less value than if it takes you to hub with greater service.

Pricing policy, such as price discrimination, is another dimension of bearer service differentiation. Essentially, it relates to the ability of a provider to efficiently allocate its network capacity according to the demand for the different levels of QoS *(best-effort, better than best-effort, service with bounded delays and packet losses etc.).* Through pricing mechanisms, bearer service providers can also target groups of consumers and charge them a unique price. This strategy allows the service provider to have lower prices in the markets where there is the greatest amount of competition while maintaining overall profitability by charging higher prices in other markets. The markets that can sustain higher prices include markets where the provider can "lock in" consumers to their service because of non-negligible switching costs or fewer competitors[21].

3.3.2. Shared residual rights of control

The brief account of bearer service product's characteristics explains, we believe, why differentiation is possible and why a bearer service market may be sustainable beyond "destructive" Bernard competition. However, this does not answer the question of whether this market is more efficient than vertical integration within the same firm "tying" in a bundled product infrastructure facilities and bearer service functionality with higher level services. S. Marble [33] seems to believe that competition is not efficient at all at the level of wholesale markets for Internet connectivity. Referring to O. Hart's theory [37], we will argue for the economic efficiency of the organizational separation between infrastructure facilities firms and providers of IP connectivity and bearer service functionality (together or not with other higher level services).

According to Hart , the critical issue for governance structure is what he calls the *rights of residual control.* These are the rights to make decisions about the usage of an asset that is not anticipated in a contract between the parties involved (incomplete contracts). There are different ways to allocate these rights, integration or non-integration, with integration involving the ownership of an asset (which goes together with the residual rights of control over the asset). The choice of the ownership structure depends upon the specifics of the relationship investments and the distribution of the information required to make decisions. Hart (op. cit) and Grossman and Hart [38] develop a theory of integration based on the above principles which allows to formally evaluate the costs and the benefits of integration (or non-integration).

To apply this framework to our case, we assume (Figure 4):

---

[21] Other ways the bearer service providers could differentiate themselves include marketing and customer service. While these differentiating factors have nothing to do with the transport of bits or the ability for an application to access the bearer service, it does influence the customers' decision when they trust their provider or achieve benefits from better customer support. Brand recognition that results from advertising and marketing can influence a decision-maker and also lead to differentiation between competing service providers. These differentiation tools are understood by the marketing literature and are beyond the scope of this chapter for adequate analysis.

i) Two different assets families, a1 and a2, infrastructure facilities (physical links and/or layer 2 switches – Frame Relay, ATM, SONET, WDM etc.) and overlying IP routers located in ISP's POPs, respectively (we consider ISP organizational structure as a network of POPs linked with Frame Relay or ATM or SONET or WDM paths).

ii) Two managers operating them, M1 and M2. M2 in combination with a2, supplies a single unit of capacity to M1. M1 uses this input to "produce" IP connectivity that is sold directly to the customers (LANs) or to other ISPs, IAPs or Service Integrators.

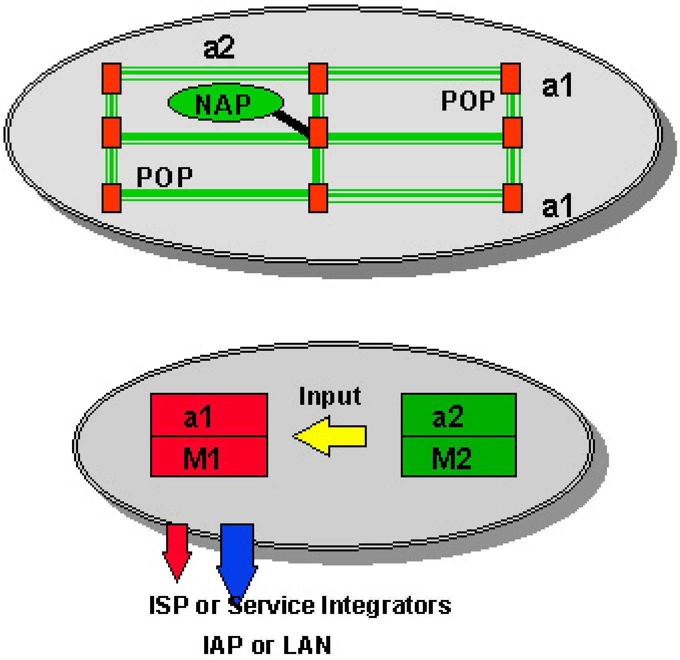


Figure 4. Assets and management decisions in the provision of bearer service

As IP packets may "run" over multiple technologies and infrastructure providers' networks, IP routers may be easily modified to work with a new underlying technology or provider network, so ISPs may switch, with no heavy costs, from one trading partner to another. Under these conditions of relative independence between a1 and a2 assets, Hart (op. cit) concludes that M1 and M2 relation-specific investments decisions should remain independent, so economic efficiency requires non-integration between M1 and M2. Independent ISPs, providing IP connectivity (or bearer service functionality) in a market, may survive and prosper.

In contrast, as long as the bearer service resided *within the wires,* the network technology defined the set of applications, so strong complementarities applied between different network assets (bitways and access points to third parties). According to Hart (ibid.), complementary assets should be under common ownership, so the rights of control over all network resources belonged to the infrastructure facilities providers (Resellers and other third parties were bound by a contract which gives them access to network resources but the infrastructure owner had the right to use the asset any way that is not inconsistent with the law. For example, the infrastructure provider would have the right to define production capacity and decide on technological investments, assuming that the initial contract is silent about these. Because the network operating firm possesses the residual rights, it receives a greater fraction of the *ex post* surplus created by its investments, so these rights work as incentives to make investments and expand the relationship with the third parties. But, at the same time, the operating firm could "veto" any allocation of the relationship assets which was not considered favorable to their interests. .

3.4. Disaggregation of the industry

What we suggest is that the impressive diffusion of the Internet may signify a transition towards flexible models of organization in the communications industry. Flexible specialization involves independence of applications' diffusion trajectories from the investment cycles of the infrastructure facilities and it significantly depends upon the existence of a bearer service interface in the "middle" of the network layered architecture. The efficient organization of a market for this bearer service is necessary for the survival of the flexible specialization: it will allow for the existence of ISPs independent from the owners of the infrastructure facilities, so able to flexibly organize their offerings based on external provisions of network capacity. The current market for IP connectivity is an early version of what we call bearer service markets with multiple levels of Quality of Service and shows opportunities for entry without leading to Bernard "destructive" price competition.

We do not claim that there will be no bearer service provider who will integrate to achieve some natural benefits. Some firms will integrate vertically to foreclose on rivals' suppliers of bearer service and other service integrators. Other firms will integrate horizontally to guarantee quality and security over an extended homogeneous network. This is already evident in the current Internet markets. However, various reasons from large *product differentiability* to limited *asset specificity* explain why *hierarchies* will exist together with *markets*.

Open layer networks are more market-friendly. To achieve economies of scale, integration of any activity within the same business organization is not *a priori* necessary. Open interfaces such as the bearer service allow *throughput* to be transferred from one layer to another, in occurrence from infrastructure facilities to upper layers. As a result, economies of scale may be obtained *externally [39,40],* through market coordination. Because of the bearer service interface, infrastructure facilities providers may spread the use of their equipment across multiple customers to obtain maximum utilization and a high level of throughput (as *Carriers of Carriers,* Qwest and Level 3 provide bandwidth to many clients including vertically integrated firms). On the other side, providers of IP connectivity and bearer service may use more than one infrastructure networks to constitute their service offerings (the new ATT-BT network is such an example), so they can flexibly reconfigure their "production" (including output levels) as a function of the demand fluctuations and the rise of new technologies and markets. When economies of scale may be obtained in those ways, externally, large vertically integrated firms are obliged to downsize and refocus their activity while new *specialized* in particular layers providers emerge and survive. That means more rich (i.e.

disaggregated) industrial structure and more market to govern the relationships between *niche* and *global* players.

References

[1] D. Clark, Interoperation, Open Interfaces, and Protocol Architecture. Laboratory for Computer Science, Massachusetts Institute of Technology, mimeo, 1995.

[2] Computer Science and Telecommunications Board (CSTB), National Research Council, Realizing the Information Future: The Internet and Beyond. National Academy Press, Washington, D.C., 1994.

[3] Computer Science and Telecommunications Board (CSTB), National Research Council, The Unpredictable Certainty: Information Infrastructure Through 2000. National Academy Press, Washington, D.C., 1996.

[4] Internet Telephony Interoperability Consortium (ITC), Workplan 1997, MIT.

[5] J. Bailey et al., The Economics of Advanced Services in an Open Communications Infrastructure: Transaction Costs, Production Costs, and Network Externalities, *Information Infrastructure and Policy* **4** (1995) 255-278.

[6] J. Gong and P. Srinagesh, The Economics of Layered Networks. In: L. McKnight and J. Bailey (eds.), Internet Economics. MIT Press, Cambridge, MA, 1997, pp. 63-77.

[7] J. Gong and P. Srinagesh, Network Competition and Industry Structure, *Industrial and Corporate Change* **5** (1996) 1231-1241.

[8] M. Piore, Corporate Reform in American Manufacturing and the Challenges to Economic Theory. Department of Economics, Massachusetts Institute of Technology, WP #90-079, 1989.

[9] M. Piore and C. Sabel, The Second Industrial Divide. Basic Books, USA, 1984.

[10] M. Piore, Beyond Individualism. Harvard University Press, Cambridge, MA, 1995.

[11] M. Piore, Fragments of a Cognitive Theory of Technological Change and Organizational Structure. In: N. Nohria and R. Eccles (eds.), Networks and Organizations: Structure, Firm and Action. Harvard Business School Press, Boston, MA, 1995.

[12] C. Kindleberger, Standards as Public, Collective and Private Goods, *Kyklos* **36** (1983) 377-396.

[13] D. Puffert, The Economics of Spatial Network Externalities and the Dynamics of Railway Gauge Standardization. Stanford University, Dissertation Thesis, 1991.

[14] D. Tennenhouse et *al*., Virtual Infrastructure: Putting Information Infrastructure on the Technology Curve. Laboratory for Computer Science, Massachusetts Institute of Technology, TM-523, 1995.

[15] G. Brock, The Telecommunications Industry: The Dynamics of Market Structure. Harvard University Press, Cambridge, MA, 1981.

[16] J. Walrand, Communication Networks. Irvin, Boston, MA, 1991.

[17] D. Bertsekas and R. Gallager, Data Networks. Prentice-Hall, Englewood Cliffs, N.J, 1992.

[18] G. Dicenet, Le RNIS: Techniques et Atoûts. Masson, Paris, 1988.

[19] M. Elton (ed.), Integrated Broadband Networks: The Public Policy Issues. Elsevier Science Pub., North Holland, 1991.

[20] R. Mansell, The New Telecommunications. Sage Pub., London, 1994.

[21] P. Kavassalis et *al.*, The Internet: A Paradigmatic Rupture in Cumulative Telecom Evolution, *Industrial and Corporate Change* **5** (1996) 1097-1126.

[22] D. Messerschmitt, The Convergence of telecommunications and computing: What are the implications today?, *Proceedings of the IEEE* **8** (1996) 1167-1185.

[23] G. Muller, International Freight Transportation. Eno Transportation Foundation, USA, 1995.

[24] P. Kavassalis and R. Solomon, Mr. Schumpeter on the Telephone: Patterns of Technical Change in the Telecommunications Industry before and after the Internet, *Communications and Strategies* **26** (1997) 371-408.

[25] D. Clark, The Design Philosophy of the DARPA Internet Protocols, *Computer Communication review* **18** (1988) 106-114.

[26] M. Rose, The Simple Book: An Introduction to Management of TCP/IP - based internets. Prentice-Hall, Englewood Cliffs, N.J, 1991.

[27] G. Dosi, Technological paradigms and technological trajectories, *Research Policy* **11** (1982) 147-162.

[28] F. Bar, Network flexibility: a new challenge for telecom policy, *Communications & Strategies* **2** (1991) 113-123.

[29] R. Braden,, L. Zhang, S. Berson, S. Herzog, and S. Jamin, Resource ReSerVation Protocol, IETF Internet Draft, February 21, 1996.

[30] S. Shenker, Service Models and Pricing Policies for an Integrated Services Internet. In: B. Kahin and J. Abbate (eds.), Standards Policy for Information Infrastructure. MIT Press, Cambridge, MA, 1995 315-337.

[31] R. Nelson, Why do Firms Differ, and How does it Matter?, *Strategic Management Journal* **12** (1991) 61-74.

[32] J. Bailey, L. McKnight and H. Sharifi, Critical Business Decisions for Internet Services, Paper presented at the Resources Management Association International Conference, Boston, May 1998.

[33] S. Marble, The Impacts of Settlement Issues on Business Evolution in the Internet, Paper presented at the 26th Telecommunications Policy Research Conference, Washington D.C, October 1998.

[34] P. Kavassalis and W. Lehr, The Flexible Specialization Path of the Internet, MIT Internet Telephony Consortium, Working Paper, 1998

[35] M. Katz, Vertical Contractual Relations. In: R. Schmalensee and R. Willig (eds), Handbook of Industrial Organization, Elsevier, North-Holland, 1989.

[36] W. B. Arthur, Increasing Returns and Path Dependence in the Economy, The University of Michigan Press, Ann Arbor, 1994.

[37] O. Hart, Firms, Contracts and Financial Structure. Oxford University Press, Oxford, 1995.

[38] S. Grossman and O. Hart, The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration, *Journal Of Political Economy* **94** (1986) 691-719.

[39] J. L. Bradach and R. G. Eccles, Price, Authority, and Trust: From Ideal Types to Plural Forms, Annual Review of Sociology **15** (1989) 97-118.

[40] T. Sturgeon, Turnkey Production Networks: A New American Model for Industrial Organization, MIT Center for Technology, Policy and Industrial Development, Working Paper, 1997.