Network tipping and the tragedy of the common network : a theory for the formation and breakdown of public telecommunications systems

Eli M. Noam
Professor, Columbia University

■ Theories for the emergence of multiple networks

According to J.K. Galbraith, "the great advantage of being in the same world as the United States is that it reveals to other countries the pleasures and horrors that will afflict them only a few years hence" (9). And while such generalization must be taken with caution, it suggests that some of recent American change in the areas of telecommunications network structure may be part of a broad and general trend of change in telecommunications, and may have relevance to circumstances as they evolve elsewhere.

Because several of the changes in telecommunications policy originated in the United States under a conservative political regime, they are often viewed as the product of particularly American business interests, wrapped in a Chicago economic ideology. But more recently, several other industrialized countries have begun to adopt similar policies, or at least to discuss changes that previously seemed unthinkable. Japan, for example, has seriously contemplated splitting its near-monopolist NTT into local exchange and long-distance companies and to give them a greater distance from equipment suppliers. Britain, Germany, and the Netherlands have lowered entry barriers. Other countries are considering similar changes. These developments raise the question whether change has explanations more fundamental than the nature of the respective governments in power. Of course, there are unique aspects to any country, and they will keep national telecommunications system to some extent distinct. But the variations should not obscure central themes that repeat themselves elsewhere.

Unfortunately, there has been little attempt at a broadergauged interpretation of the transformation of telecommunications networks that can explain the dynamics of change (10). To provide such wide analysis is the aim of this essay.

---

(9) GALBRAITH, Interview with Francis Cairncross, The Observer, 22 November 1970.
(10) One attempt was a US Department of Justice report on the post-divestiture network : HUBER Peter, *The Geodesic Network*, Washington DC : US Goverment Printing Office, 1987. Another approach is that of HAYASHI Koichiro of NTT, *The Economies of Networking. Implications for Telecommunications Liberalization*, paper presented at the ICC Conference, Washington, D.C., Sept. 1988. An earlier version of the present model was : NOAM Eli, *The Next Stage in Telecommunications Evolution : the Pluralistic Network*, paper presented at the ITS Conference, June 1988.

A number of explanations have been offered -explicitly or implicitly- for the demise of monopoly in telecommunications. There are three major types of theories.

Technological explanations

This perspective comes in two variants :
(A) "More powerful technology leads to new transmission options and thereby to competition and the breakdown of monopoly".
(B) "The merging of telecommunications and computing technologies breaks down traditional barriers separating different industries and undermines monopoly power".

These views are typically held by technologists, and they are influential in an engineering-oriented industry such as telecommunications. But they are not sufficient as explanations, or else one would observe a diversity of physical networks also emerging in, say, France, Australia, or Mexico. After all, the same transmission and switching technologies are available anywhere on the globe (11). Yet their impact on network structure has varied, and provide no evidence for a technological determinism at work. Technology change provides the precondition for change, but it is not a sufficient condition.

Political explanations

Here, too, there are two related versions, both using the perspective of countervailing powers :
(C) "In the information age, a telecommunications monopoly becomes too powerful and its scope needs to be limited".
(D) "Government regulation proves incapable of controlling a monopoly, and is therefore replaced by policies encouraging a competitive industry structure".

The problem with these views is that the introduction of a multiplicity of carriers is only one policy option out of several. An alternative response to political power or regulatory inefficiency might well be a stricter or more effective regulation, as would be nationalization or a size-reduction along geographical and/or functional lines while maintaining monopoly. Thus, it is not clear why competition is the necessary remedy to monopoly power.

Non-sustainability explanations

Another view is that a monopoly, even if efficient across its multiple products, cannot protect itself from entry into some lines of business.

(E) "The diversification of telecommunications makes it difficult for any one provider to serve all sub-markets without competitive entry".

(11) Exceptions are CoCom restrictions to East bloc countries which are fading, with the latter.

This view is essentially that of an economic non-sustainability theory advanced by Baumol et al.(12). It can explain the emergence of entrants for new products of a multi-product firm, but it does not adequately cover competition in traditional core markets of a telecommunications monopolist, unless one accepts very restrictive assumptions (13).

Market structure explanations

There are two variants to this approach, one passive and the other active :
(F) "Monopoly's inefficiency eventually leads to the emergence of competition".
(G) "Competition is a policy chosen to enhance efficiency and technological development".

These views are held by many economists. Theory (F) is a Milton Friedman type view of the impermanence of privileged economic arrangements. Theory (G) expects governments not to simply wait for competitors' entry but to institute proactive and pro-competitive policies. Examples are the United Kingdom and Japan, where competition was introduced from above. These two views have in common the premise of inefficiency of monopoly. In other words, a multi-carrier market structure is believed to be emerging due to some failure of the traditional system. Yet this assumption is at tension with the reality of network performance in those countries where structural changes in networks is most rapid. If inefficiency were the causal force for rival entry, Egypt or Mexico (to use two examples) should have introduced competition long before the US and Japan, which had arguably the most advanced and ubiquitous networks in the world even before embarking on their liberalizing policies.

It has always exasperated the proponents of the traditional network system to be told that their problem was inefficiency. This clashed with their observations of economies of scale, benefits of long-term technological planning, and effectiveness of end-to-end responsiblity. Thus, explanations based on the inadequacies of the monopoly system are not persuasive.

A new approach : the tipping of network coalitions

None of these seven theories for the emergence of multiple networks provides an adequate explanation, though they all contain some truth, and their aggregate holds some explanatory power.

In contrast, this essay advances an eighth and alternative view, that of the dynamics of group formation. The thesis of this essay is not based on the failure of the existing system but on its success. Changes in technology politics, and cost are merely enabling a more fundamental shift of coalitions.

(12) BAUMOL William J., PANZAR John, C. and WILLIG Robert D., *Contestable markets and the theory of industry structure*, New York, Harcourt Brace Jovanovich, 1982.
(13) SHEPHERD William, *Concepts of competition and efficient policy in the telecommunications sector*, in Eli M. NOAM ed., Telecommunications today and tomorrow, New York, Harcourt Brace Jovanovich, 1983.

(H) "The breakdown of monopoly is due to the very success of the traditional system in advancing telephone service and making it universal and essential. As the system expands political group dynamics take place, which lead to redistribution and overexpansion. This provides increasing incentives to exit from a sharing coalition, and to an eventual "tipping" of the network from a stable single coalition to a system of separate sub-coalitions".

This view of success undermining its own foundations is basically Schumpeterian. From the monopoly's perspective, it is deeply pessimistic, because it implies that the harder their efforts and the greater their success, the closer the end to their special status is at hand. Like in a Greek tragedy, their preventive actions only assure their doom.

■ The concept of network

Networks are an important concept in society and the economy. They abound as physical facilities, such as those of electric utilities, communications, and transportation. Networks are also relational systems, such as those of "old boys", political supporters, and intelligence agents.

The term "network" goes a long way back ; it is used, in the King James translation, by the Supreme Regulator : "And the Lord spake unto Moses, saying... You shall also make it a grating, a network of brass..." Exodus XXVII, V. 4. In Hebrew, the word is "reshet" (net) similarly used today for telecommunications and other networks.

The term is used by most academic disciplines, and with a variety of meanings. Chemists apply it to arrangements of molecules (14). Biologists to cell structures (15). Mathematicians to topology (16). Electrical engineers to distribution systems (for high voltage), or for circuit elements (for weak voltage) (17). Their network theory deals with capacitance, inductance, etc., for systems of components.

Operations researchers use a network terminology to solve shortest path problems, maximum flow models, and optimal routing (18). Computer scientists apply the term for computer interconnections in hardware, and to implementation algorithms in software (19).

In the social sciences, political scientists use the concept of networks in discussing hierarchies, interactions, gatekeepers, and policy communities (20). For sociologists and social anthropologists, networks are a major way to see the world ; a basic point is that the nature of linkage affects behavior (21). Sociologists speak of network dyads -interpersonal linkage between two persons in which each is indebted to the other, and similar in some ways to the exchange the relation of economics.

---

(14) ZACHARISEN, W. M., *The atomic arrangement in glass*, October 5, 1932, in Journal of the American Chemical Society, Vol. 54, n° 10, p. 38-42, Washington D.C.
(15) KNOX Robert, *Elements of general anatomy (translated from Inst. edition of Beclard's Anatomy,* by D.A. Beclard, Edinburgh, Scotland : Maclachlan & Stewart, p. 214, 1830.
(16) KLINGMAN David J. and MULVEY J. eds., *Network models and associated applications*, Amsterdam and New York, Elsevier North-Holland, 1981.
(17) KARNI Shlomo, An analysis of electrical networks, New York, Wiley, 1986, pp. 1-4.
(18) ELMAGHRABY Salah E., *Some network models in management science*, New York, Springer, 1970, pp. 1-3.
(19) KLINGMAN David, J. and MULVEY J. eds., *Network models and associated applications*, Amsterdam and New York, Elsevier North-Holland, 1981.

Among the social science disciplines, economists have probably paid the least attention to networks. There is no body of analysis for the network concept. Somewhat related is work on market structure by some industrial organization theorists (22). Closer are public choice theories of group formation, discussed in the following section. Other writings on networks are Noam (1988) and Heal (1989) (23).

Corresponding to different disciplines' use of the network concept, they also approach concrete applications in different ways. Thus, when it comes to telecommunications networks and network policy, several ways of thinking can be distinguished. They are the golden calves worshipped by different professional denominations.

For technologists a primary organizing concept are the protection of technical efficiencies. Economists, on the other hand, tend to worship at the altar of competition, mostly to the triad of structure, conduct, and performance. And while there is increasing disenchantment with this view, it is more represented in academia than in the regulatory environment.

Thirdly, lawyers in this field judge often policy issues in terms of "conflict of interest", which translates here to potential for cross-subsidies. This perspective is particularly developed in the United States. Hence, the AT&T divestiture, and the proceedings of Computer I, II and III.

In contrast, many other social scientists as well as most politicians and journalists organize reality in telecommunications policy around the concept of income distributions, that is, around the question who pays and who receives, and what factors of political power lead to such distribution.

All of these concepts have legitimacy, but have been carried by their proponents beyond their explanatory power. Used singlemindedly, they have degenerated from tools of analysis to rallying slogans in policy disputes.

---

(20) RICHARDSON, Jeremy John and GUSTAFSON Gunnel and ART Jordan, as cited in RHODES, R.A.W., *Power dependence, policy communities and intergovernmental networks*, Colchester, Essex, Department of Government, University of Essex, Wivenhoe Park, 1985, pp. 6-8.
(21) - BARNES, J.A., *Class and committees in a Norwegian Island Parish*, in Human relations, vol. 7, LAZARSFELD Paul and MERTON, Robert, 1954.
- BOTT E.., *Family and social network*, London, 2nd ed., 1971.
- BOISSEVAIN J., *Network analysis : a reappraisal*, in Current anthropology, vol. 20, 1979.
(22) BAUMOL William, PANZAR John and WILLIG Robert, *Contestable markets and the theory of industry structure*, Harcourt Brace Jovanovich, New York, 1982.
(23) NOAM Eli, *The next stage in telecommunications evolution : the pluralistic network*, paper presented at the ITS Conference, MIT, June 1988, working paper # 316, Columbia University CTIS.
- HEAL Geoffrey, *The economics of networks*, Columbia University, unpublished paper, 1989.

Perhaps the greatest common failing of these traditional organizing ways of looking at telecommunications principles is that they concentrate on "supply-side" analysis. That is, they look at the subject from the angle of production and producers : AT&T v. MCI, Intelsat v. Cable & Wireless, value added v. basic carriers, etc.

It is not surprising that this approach would be taken. After all, policy makers deal primarily with carriers, technologists with networks, economists with competitors, and journalists love a horse-race angle to their coverage. But this supply-oriented perspective obscures the other side of the coin - what could be called a demand-side telecommunications analysis. At its most basic, one should not think of telecommunications primarily as a service produced by carriers, but as an inter-action of groups and subgroups in society, facilitated by service suppliers called carriers. The supply structure, if left to its own devices, is a reflection of the underlying interaction of communication users with each other, whether in one all-encompassing "user coalition", or in several user groupings. A universal public network that interconnects everybody with anybody, under a single organizational roof, is technically and financially merely one arrangement out of several.

Thus, one should not view deregulation as a policy of primarily liberalizing the entry of suppliers. Just as importantly, though much less obviously, it is the liberalization of exist, by some partners, from a previously existing "sharing coalition" of users which has become confining.

Integration and centrifugalism are two basic types of forces -call them communalism vs. particularism, order vs. diversity common to many social processes. Telecommunications are only one instance for the wide-spread ascendancy, in recent years, of centrifugalism in previously shared arrangements. Wherever one looks, people break up all kinds of social networks of interaction and form new ones.

Examples abound. They include, for the United States, public education, mass transit, public safety and dispute resolution, pension and health provision, electrical power and gas distribution, stock exchanges, etc.

The departure, in the United States, of many participants from the public school system is not explained primarily by the supply of new options or by new technology, but rather by increased demand to exit. Similarly, recent emergence of "independent" electric power generation has little to do with new technology.

It would go beyond the scope of this essay to discuss the factors underlying these transformations. But focusing on telecommunications, we can look at its key institutional relation, the shared network, and analyze why its unity breaks down.

■ A model of networks

Perhaps the best way to look at a network is as a cost sharing arrangement between several users. Fixed costs are high, marginal costs low, and a new participant C helps the incumbents A and B to lower their cost. In that it is similar to a swimming pool or national defense, that is, to a "public good". But while there is basically only one national defense system, there are many types of arrangements for swimming pools. A user may want to share the pool with a few dozen families, but not with thousands. A pure public good admits everyone, a pure private good, only one. But there is a wide spectrum between the pure private good and the pure public good" (24). A telecommunications network is one intermediate example. It is not a private good, yet it does not meet the two main conditions for a public good : non-rival consumptions and non-excludability. In fact, non-excludability has to be established as a legal requirement -the universal service obligation. What has been happening in recent years to telecommunications, and what goes by the more dramatic labels of divestiture and deregulation, is largely a shift in the degree of its intermediate position, a shift toward the direction of private good.

We will now develop, in a stepwise fashion, a model for network evolution and diversification.

The basic model (25)

Let the total cost of a network serving n subscribers be given by :

$TC(n) = F + f(n)$ (1)

where fixed cost $F \geq 0$ and marginal cost $f'(n) \geq 0$.

n assumes that users are homogenous. Of course, some network participants are much larger than others, but that poses no problem if we define a large organization to consist of multiple members of type n, e.g., telephone lines or terminals rather than subscriptions. Later, we will drop this assumption. Let an individual's utility be given by u (P, n), where P is the price for network usage, and n are the number of network members (26). We assume network externalities to exist, ($\partial u/\partial n > 0$), though at a declining rate ($\partial 2 u/\partial n2 = < 0$).

i.e. a subscriber is better off the more other members there are on the network, *ceteris paribus* (including network performance and price) (27). For simplicity, utility is expressed in monetary units.

$u = u(P) + u(n) = P(n) + u(n)$ (2)

We assume that the network membership is priced at average cost, i.e. that users share costs equally. (This assumption will be dropped later). This can be

(24) BUCHANAN James M., *An economic theory of clubs*, Economica, 32, n° 125, 1965, pp. 1-14.

(25) I will follow the network analysis as developed in Noam Eli, *The next stage in telecommunications evolution : the pluralistic network*, paper presented at the Pacific Telecommunications Conference, Japan, October 1988, and begin in sections 1-4 with some of the methodology of my Columbia colleague Geoffrey Heal, *The economics of networks*, Columbia University, unpublished paper, 1989, though with significant differences.

(26) Strictly speaking, income is allocated to telephone and to other consumption :
$y = cp + (F + f(n))/n$.
A subscriber's utility is then given by :
$u(y/p - (F + f(n))/np; n)$.

(27) For convexity, assume u (c,P,1) > u (c, P,o), i.e. the first user has positive benefits even if no one else is on the network.

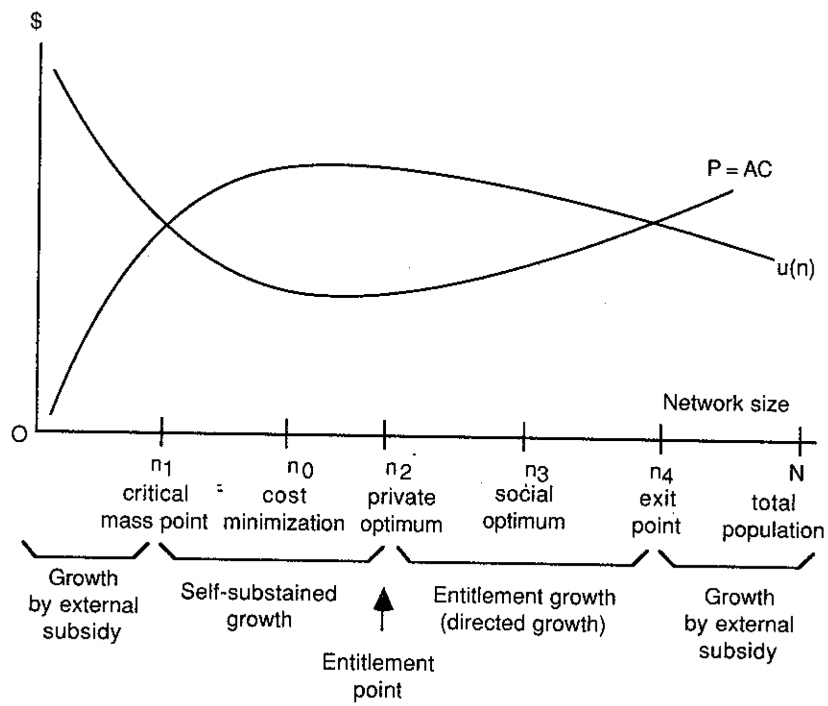
shown schematically in graph 1, where u(n) is steadily increasing, though at a declining rate, and $P = AC = [F + \partial (n)]/n$ is declining, at least at first. The network, at this stage, is in its cost-sharing phase.

Critical mass

Subscribers will find it attractive to join a well-sized network, because total costs are shared by many, making average costs low, while the number of subscribers n adds to utility. This can be seen in Figure 1, where utility of joining a network rises at first. Conversely, where the network is small, average cost is high, and externalities small. In that range, below a "critical mass" point n, a network will not be feasible, unless supported by external sources. We define critical mass to be the smallest number of users such that a user is as well off as a non-user $u(n) = P(n)$ (28).

To reach n, requires a subsidy of sorts, either by government or by the network operator's willingness to accept losses in the early growth phases of operations. The strategic problem is to identify in advance a situation in which such a break-even point n, will be reached within the range $n < N$, where N = total population. Possibly, such a point does not exist, and subsidies would have to be permanent in order to keep the network from imploding. We will return to the critical mass issue later in subsection 11.

Figure 1



(28) Heal defines it, similarly, as : u (y/p - [F = f(n)]/hp; n) = u (y/p; 0).

Private optimum

Through the cost-sharing phases of network growth, the earlier network users can lower their cost by adding members. However, at some point average cost AC increases.

$$\frac{dAC}{dn} = -\frac{F}{n^2} - \frac{f(n)}{n^2} + \frac{f'(n)}{n} = \frac{1}{n}\left[\frac{F + f(n)}{n} - f'(n)\right] = -\frac{1}{n}[AC - f'(n)]$$

That is, average cost increases in the range beyond the point n_0 where $AC = f'(n)$.

Beyond n_0 expansion becomes unattractive for cost reasons. But, some further expansion would be accepted by the network members since newcomers beyond the low cost point would still add to utility. This will be up to the point n_2 where the total derivative with respect to the number of users is equal to zero.

$$\frac{du}{dn} = \frac{\partial u}{\partial c}\frac{\partial AC}{\partial n}\frac{1}{p} + \frac{\partial u}{\partial n} = 0 \qquad (4)$$

which holds where the marginal utility of an added subscriber is :

$$\frac{\partial u}{\partial n} = \frac{\partial u}{\partial c} \cdot \frac{1}{p}\frac{\partial AC}{\partial n} \qquad (5)$$

This is the case in the range of increasing AC. Graphically, n_2 would lie where the two derivatives are of equal size, u' (P) = u' (n). Left to themselves, the existing subscribers of the network would not accept members beyond n_2, the private optimum.

Social optimum

From a societal point of view, however, the optimal network size in an equal price system may diverge from the private optimum.

Assume social welfare given by the sums of utilities.

$$W = n\,[u(P(n)) + u(n)] = n\,(F + f(n))/n + u(n)] \qquad (6)$$

so that :

$$\frac{dW}{dn} = f'(n) + n\,u'(n) + u(n) = 0 \qquad (7)$$

Social welfare increases at n_2, and becomes zero at a point of intersection n_4. Hence, social optimum n_3 is somewhere in between those two points.

What is the implication ? Left to itself, the network association will cease growth beyond n_2, at least as long as costs are equally shared. Existing network subscribers would not want to admit newcomers beyond n_2. Latecomers beyond that point add cost, because they raise AC, and add fewer externality benefits. In the words of Alfred Kahn : "People who don't have a phone I don't want to talk to". The socially optimal size n_3 will not be reached by itself, but through some external governmental direction through required expansion, and/or a differentiated pricing scheme, or through some internal politics of expansion that will be described below.

Entitlement point and universal service obligation

To understand the politics of government-directed network expansion, let us assume a political decision mechanism in which majorities rule. When private optimum size $n_2 < N/2$, there are more people outside than inside the network, while there are positive net benefits, i.e., $u(n_2) - AC(n_2) > 0$. A majority consisting of $N - n_2$ network outsiders would therefore outvote the n_2 network insiders, and require the opening of the network to additional members. This would be the case up to the point where network size reaches N/2, at which point the network insiders have grown to a majority and will resist further growth. Beyond N/2 then (or where $n_2 \geq N/2$ and a majority against expansion exists from the beginning) a politically directed growth will occur if the coalition of network insiders can be split by aligning the remaining outsiders N/2 with some of the insiders who are offered a more favorable share of cost, i.e., by price-discrimination. It can be shown that this coalition formation will lead to an over-expansion of the network. The dynamics of such price discrimination and its impact are discussed in greater detail in this chapter's Sections 7 and 8 that deal with expansion and network tipping.

Politically directed growth beyond private optimum n_2 can be termed an "entitlement growth" because it is based on political arguments of rights to participate in the network where average net benefits are positive (encouraging attempts of entry) while marginal net average benefict are negative, leading to attempts at exclusion. When the marginal net benefits are positive, there is no need to resort to the language of entitlements, since growth is self-sustaining and sought by network insiders. It is only beyond that point that entitlements, rights, and universal service rights (i.e. obligations by the network) become an issue. We can thus define n_2 as the "entitlement point".

This way of analyzing entitlements serves to clarify the often-asked question : for which services will universal service be extended ? Using the analysis, the answer is to those services that :

(a) have grown beyond minimum critical mass and

(b) have reached, through self-sustained growth, a private optimum, beyond which further growth is not internally generated because marginal average net benefits are zero, but where

(c) average net benefits are positive (and therefore encourage demand for entry), and

(d) the number of those excluded is sufficiently large to lead to an opening by political means.

Exit from the network

If $n_2 < N$, with N being the total population, some people would be left out of the network. But as discussed in the previous section, a government would require for the network to be open to other users. Yet there may well be a point where the network is expanded to an extent that, given its increasing

marginal cost, a user is better off by not participating. We define n_4 as the "exit point", i.e., the largest n such that the indifference exists between dropping off the network and sharing in the cost of supporting the expanded network.

$$u(n) = u(P) \qquad (8)$$

It is possible that this exit point lies beyond the total population, $n_4 > N$. But this seems not likely under an average-pricing scheme, because the last subscribers may impose a heavy burden on the rest of subscribers, and the subsequent departure of some subscribers would lead to further reduction in the utility of the remaining members and may induce a secondary exodus. Thus, assuming $n_4 < N$, a government's aim to establish a truly universal service, without resorting to a subsidy mechanism or price discrimination, is likely to be infeasible. In other words, a universal service policy is likely to be dependent on a redistributive policy.

Political price setting and redistribution

We have so far assumed that universal service is something imposed externaly by government. In this section, however, it will be shown that the internal dynamics of network members will take the network towards universal service, and towards its own disintegration.

As has been shown, a network will cease to grow on its own after private optimum n_2. But this conclusion was based on a pricing scheme of equal cost shares. Yet there is no reason why such equality of cost shares would persist if they are allocated through a decision mechanism that permits the majority of network users to impose higher cost shares on the minority. (This assumes that no arbitrage is possible). Unequal prices and a departure from cost could be rationalized benignly as merely "value of service" pricing, i.e. higher prices for the users who value telephone greatly.

Suppose for purposes of the model that decisions are made through voting by all network members (29). Let us assume as this stage that all users are of equal size (or that voting takes place according to the number of lines a subscriber uses, which is the same thing) and that early network users have lower demand elasticity for network use. The determinative vote is provided byt he median voter located at n/2. A majority would not wish to have its benefits diluted by a number of beneficiaries larger than necessary. This is the principle of the "minimal winning coalition". Its size would be n/2 + 1.

A majority will establish itself such that it will benefit maximally from the minority. The minority that can be maximally burdened are the users with less elastic demand for telephone service, which are the early subscribers. But there is a limit to the burden, given by utility curve u(n). If price gets pushed above u(n), subscribers will drop off. Hence, the majority

($n_2/2 + 1$) will burden the minority

---

(29) This analysis should not suggest that a voting mechanism is governing in reality (although it exists for telephone cooperatives in Finland and the US) but rather to understand the pressures and dynamics that are transmitted to the governmental institutions which embody the different user interests.

($n_2/2$ - 1) with a price up to positive utility, and they will bear the rest of the cost. The minority's price P_1 will be such that :

$$P_1 = u(n_2). \qquad (9)$$

The majority's price will then be (30)

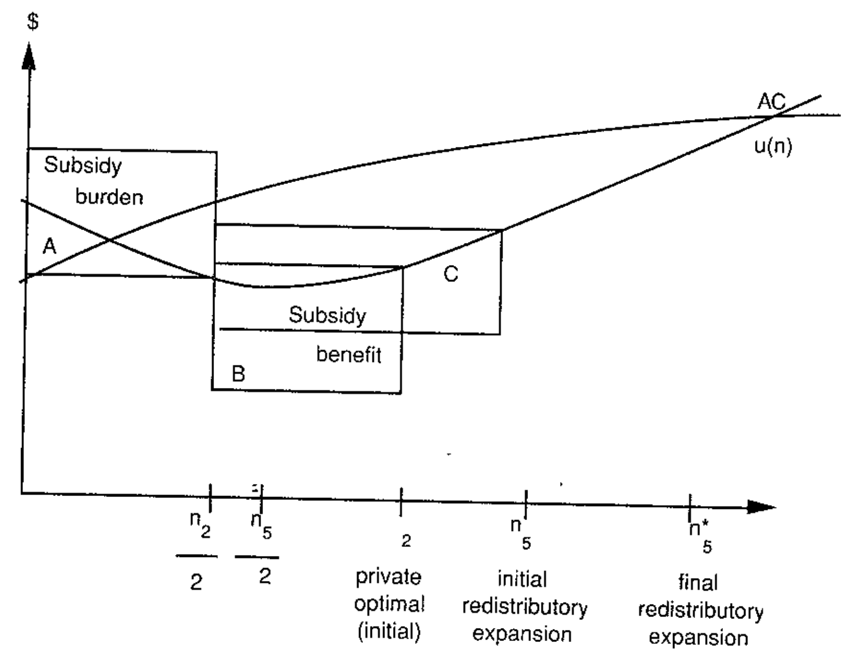
$$P_2 = (F + f(n_2) - n_2/2 \,.\, P_1)/(n_2/2 = 2AC - P_1 \qquad (10)$$

Graph 2 shows this relation in a different way. Suppose u(n) gives a subscribers's utility for various network sizes n, expressed in monetary units, and P = AC(n) is the price vector. Private optimum is at n_2. The minority $n_2/2$ can be charged a price P_1, where P is equal to utility $u(n_2)$. The shaded area constitutes the subsidy ; B is total cost ; remaining majority is C = B - A, and the price charged to majority subscribers is :

$$C/(n_2/2), \text{ or } P_2 = [F + f(n_2) - (n_2/2)\, u(n_2)]/(n_2/2) \qquad (11)$$

$$P_2 = 2AC - u(n_2) = 2AC - P_1 \qquad (12)$$

This then is the redistributory outcome, assuming no discrimination within majority and minority.



(30) For simplicity, we use in the following n/2 rather than (n/2) - 1.

Monopoly and expansion

But such redistribution is not a stable equilibrium. Before, network size n_2 was reached (once the critical mass threshold was crossed) by voluntary association. Further members were not admitted, because they lowered utility to the incumbents. But with internal redistribution, several things happen. There are now incentives for the minority network members to exit the network and form a new one in which they would not bear the redistributory burden. This would be possible if the minority were of a size larger than critical mass, $n_2/2 > n_1$. Even where that were not the case, the minority could band together with those beyond network size n_2 who desire telephone service but were previously excluded.

This exit would deprive the majority of the source of its subsidy and is therefore held undesirable. The only way for the majority to prevent this "cream-skimming" or "cherry-picking" is to prohibit the establishment of another network, both by those wanting to leave the original network and similarly by those not admitted to it by being beyond n_2.

Thus, a monopoly system and the prevention of arbitrage become essential to the stability of the system.

At the same time, and importantly, the model predicts that the network must expand beyond n_2. For the majority, there is added utility from added network members, while most of its cost is borne by the minority. They will therefore seek expansion. The cost to the majority is only that the subsidy by the minority must be shared with more network participants. Therefore, the majority would admit new members up to the point n_5 where marginal utility to its members is equal to the marginal price due to the diluted subsidy.

But this is not the end of the story ! With expansion to n_5, the majority is now $n_5/2$ rather than $n_2/2$, i.e. larger than before, and it can also tax a larger minority (ns/2) than before. Hence, the expansion process would take place again, leading to a point $n'_5 > n_5$. This process would continue, until an equilibrium would be reached at the point where for a minority member :

$$\frac{du}{dn} = o, \text{ with} \qquad (13)$$

$$\frac{du}{dn} = \frac{\partial u}{\partial n} + \frac{\partial P2}{\partial n} \cdot \frac{\partial u}{\partial P_2} \qquad (14)$$

$\partial u/\partial P = 1$ by assumption that utility is expressed in monetary units.

P_2 is given by $u(n_2) - [F + f(n_2)]/n_2$, and therefore :

$$\frac{\partial P_2}{\partial n} = \frac{\partial u}{\partial n} + \frac{F}{n_2} - n\,\frac{f'(n)}{n_2} \pm \underline{f(n)} \qquad (15)$$

So that :

$$\frac{du}{dn} = \frac{\partial u}{\partial n} + \frac{\partial u}{\partial n} + \frac{F + f(n)}{n^2} - \frac{f'(n)}{n} \qquad (16)$$

$$\frac{du}{dn} = \frac{2\partial u}{\partial n} + \frac{1}{n}(AC - MC) \qquad (17)$$

The optimizing point n_5 is where du/dn = o, i.e. where the equation holds :

$$\frac{\partial u}{\partial n} = (MC - AC)/2n \qquad (18)$$

This is a remarkably simple result. n_7 is the point up to which the network will grow under the internal dynamics described above. It will be larger, the greater the marginal utility from added network member is, the smaller marginal cost, and the greater fixed cost are (31).

Network tipping

As this process of expansion takes place, the minority is growing, too. The likelihood that its size increases beyond the point of critical mass n_1 is increased, and the utility of its members, given the burden of subsidy, may well be below that of membership in a smaller but non-subsidizing alternative network. We have so far assumed that there is only one network, and that a user's choice is whether to join or not. Suppose there are no legal barriers to the formation of a new network. In that case, a user's choice menu is to stay, to drop off altogether, or to join a new network association. Assume that the new network would have the same cost characteristics as the traditional network has. (In fact, it may well have a lower cost function for each given size if there has been accumulated monopolistic inefficiency in the existing network and rent-seeking behavior by various associated groups).

Then, minority coalition members would find themselves to be better off in a new network B, and they would consider such a network, abandoning the old one. The only problem is that of transition discontinuity. A new network, in its early phases, would be a money-losing proposition up to its critical mass point n'_1 (32).

The point where exit becomes possible, given the redistributory burden that keeps utility just balanced with price, is the point n beyond where :

$$u(n/2) > AC(n/2) \qquad (19)$$

which is at the critical mass point. Hence, the majority will strive to exit the redistibutory network once the latter's size is more than twice the size of critical mass.

The majority may attempt to alleviate these pressures to exit by reducing the redistributory burden and thus keeping the minority from dropping out. But that means the network size n_7 would not be optimal to the majority anymore, and members would have to be forced out. And this, in turn, would

(31) A very similar analysis can be undertaken for a political system in which network-outsiders have a vote, as in a governmentally directed system.
(32) This corresponds to reality ; new networks such as MCI and Sprint have lost money for a number of years until turning profitable with growth.

reduce its majority, so that it would have to drop the subsidizing burden from at least some minority members as the n/2 point separating the majority from the minority shifts leftwards.

This means either higher burdens on the shrinking minority -frustrating the purpose of bribing it into staying- or still less benefits for the majority if it wants to keep the network from fragmenting. Such a disequilibrium process will continue up to the point where network size $n = 2\,n_1$, i.e. where the minority may be too small to create a self-supporting new network. One might call this the effect of potential exit by the minority, and results in a lessened redistribution of newcomers to the network in order to keep the first entrants inside.

Unequal user size

We have assumed so far the network voters are of equal size. In reality however, some users are much larger in terms of lines n than others. The minority's position would be further weakened if voting were governed by a principle of "one subscriber, one vote" rather than the "one line, one vote" previously assumed.

Suppose users are ordered according to size on Figure 1 ; in other words, the largest users are those that have joined the network first. This is not unrealistic, since users with great needs for telecommunications are likely to have been the first to acquire a telephone, and early subscribers had the longest time to expand usage. Let us further represent the distribution of lines n for a user v by $n = Av^a$ (20)

where $A > o$, $a \geq 1$

The median voter (or median account) is v/2 and its preferences govern. But the network size provided by the users arrayed to the left of such median user is larger than those to the right. They are given by :

$$n_m = A \int_{n/2}^{n} v^a \, dv = (A/(1-a))\,(1 - n/2^{1-a}) \qquad (21)$$

n_m, the median account, is to the right of n/2 in Graph 1. In other words, the median voter whose preferences govern is at a network size greater than the median point of the network size. The more the distribution of lines is skewed (the larger the coefficients A and a) the further to the right is n_m. And the more skewed the distribution, the more likely is it that the voting minority will reach, by itself, a size beyond the critical mass point.

Interconnection

The process of unravelling of the existing network would commence even earlier if a new network has the right to interconnect into the previous one, because in that case it would enjoy the externality benefits of a larger reach $n_A + n_S$, while not being subject to redistributory burden.

Would there exist, for any sub-network, internal redistribution based on coalitions ? Once the possibility of exit is established, each burdened sub-group could join another network. Thus, internal redistribution will happen only if a network is unique to some extent, and thus if a burdened user will not readily switch into another network.

Network interconnection means that the network still centers around as a society-wide concept of interconnected users. But it consists now of multiple subnetworks that are linked to each other. Each of these subnetworks has its own cost-sharing arrangements, with some mutual interconnection charges. Interconnection facilitates the emergence of new networks. It lowers entry barriers. But given entry, it may reduce competition by establishing cooperative linkages instead of end-to-end rivalry (33). But as the next section will show, it also may lead to market failure in the establishment of the original network.

Subsidies for reaching critical mass

We have mentioned before that waiting for demand to materialize prior to the introduction of a network or network service may not be the optimal private or public network policy. Demand is a function of price and benefits, both of which are in turn functions of the size of the network. Hence, early development of a network may require internal or external support in order to reach critical mass.

This suggests the need, in some circumstances, to subsidize the early stages of the network -up to the critical mass point n_1-when the user externalities are still low but cost shares high. These subsidies could come either from the network provider or its membership as a start-up investment, or from an external source such as a government as an investment in "infrastructure", a concept centered around externalities. The question is how the internal support is affected by the emergence of a system of multiple networks.

The private start-up investment in a new form of network is predicated on an expectation of eventual break-even and subsequent positive net benefits to members. But if one can expect the establishment of additional networks, which would keep network size close to n_1, there would be only small (or no) net benefits realized by the initial entrants to offset their earlier investment. This would be further aggravated by interconnection rights, because a new network could make immediate use of the positive network externalities of the membership of the existing network that were achieved by the latter's investment. Hence, it is less likely that the initial risk would be undertaken if a loss were entirely borne by the initial network participants while the benefits would be shared with other entrants who would be able to interconnect and thus immediately gain the externality benefits of the existing network users, but without contributing to their cost-sharing. The implication is that in an environment of multiple networks which can interconnect, less start-up investment would be undertaken . It pays to be second. A situation of market failure exists.

(33) MUELLER, MILTON, *Open interconnection and the economics of networks : an analysis and critique of current Policy,* paper presented at Telecommunications Policy Research Conference, Airlie, Virginia, Oct. 31 - Nov. 1, 1988.

How could one offset this tendency if it is deemed undesirable ? Patents are one solution. If a contemplated new network arrangement is technologically innovative, it might obtain a patent protection for some period. Where a service is innovative but not patentable, one might create a "regulatory patent" for a limited period of protection, or the initial approval (where necessary) might be accelerated. Similarly, interconnection rights might be deferred for a period, or joint introductions be planned that eliminate the first entrant penalty. But these measures would also reduce the usefulness of alternative networks, and could hence lead to the dynamics of political expansion, redistribution and break-up described in earlier section.

It is possible, moreover, that none of these measures would be as effective in generating the investment support in the way that a monopoly network would that can reap all future benefits. This would mean that the private and social benefits of networks in the range between n_1 and n_4 would not be realized. In such a situation, there may be a role for direct outside support, such as by a government subsidy. This may strike one at first as paradoxical. Shouldn't a competitive system of multiple networks be less in need of government involvement than a monopoly ? But on second though, there is some economic logic to this. Just as the subsidies to individual network users that were previously internally generated by other network users will have to be raised externally (through the normal mechanism of taxation and allocation) if at least some users are still to be supported, so might subsidies to the start-up of a network as a whole have to be provided externally, also through taxation and allocation, where network externalities as well as start-up costs are high enough to make the establishment of a network desirable.

Social welfare and multiple networks

If network associations can control their memberships, stratification is inevitable. They will seek those members who will provide them with the greatest externality benefits -those that have many actual or potential contacts with. Furthermore, they will want to admit low-cost, high volume, good risk customers as club members. Thus, different affinity-group networks and different average costs will emerge.

But what about social welfare in such a differentiated system ? The traditional fear is that the loss of some cost-sharing and externalities brought by a second network would reduce social welfare. But the news is not necessarily bad. Where the network was at n_3 or substantially larger than the socially optimal size n_4, the fracture of the network could increase social welfare, depending on the cost and utility functions, if cost closer to n_0 is reached. Where mutual interconnection is assured, one can keep the externalities benefits (and even increase them) while moving down the cost curve towards a lower AC. Furthermore, the cost curves themselves are likely to come down with the ensuing competition.

The welfare implications of the formation of collective consumption and production arrangements is something analyzed by theorists of clubs.

(Schelling, Buchanan, Tullock, Rothenberg, Tiebout, and McGuire) (34). The club analysis, applied to networks, can show :

1. Given mobility of choice, different groups will cluster together in different associations according to quality, size, price, interaction, and ease of internal decision-making. The economically optimal association size need not encompass the entire population (35).

Optimal group size will vary according to the dimension to be optimized. Optimal group size depends on the ratio of marginal utilities for different dimensions, set equal to the ratio of transformation in production, and is in turn related to size (Buchanan, 1965, p. 4-5) (36).

But this does not imply that one should keep networks non-ubiquitous and unequal. Financial transfers can be used.

However :
2. It is generally not Pareto-efficient to attempt income transfer by integrating diverse groups and imposing varying cost shares according to some equity criteria. It is more efficient to allow sub-groups to form their own associations and then redistribute by imposing charges on some groups and distribute to others. The set of possible utility distributions among separate groups dominates (weakly) the set of such distributions among integrated groups (McGuire; JPE, p. 124). User group separation with direct transfer is more efficient than the indirect method of enforced togetherness with different cost shares. In other words, differentiated networks plus taxation of another system of revenue shifting such as access and interconnection charges, is more efficient than monopoly and internal redistribution.

Conclusion

The theoretically-based analysis of the model means that a network coalition, left to itself under majority-rule principles, would expand beyond the size that would hold under rules of equal treatment of each subscriber. Such an arrangement can be stable only as long as arbitrage is prevented, as long as the minority cannot exercise political power in other ways, and most importantly, as long as it has no choice but to stay within the burdensome network arrangement.

---

(34) SCHELLING Thomas C., *Models of segregation*, Santa Monica, Rand, 1969.
- BUCHANAN, op. cit.
- TULLOCK, GORDON, *Public decisions as public goods*, Journal of Political Economy, n° 179, n° 4 : 913-918, July-Aug., 1971.
- ROTHENBERG Jerome, *Inadvertent Distributional impacts in the provision of public services to individuals*, in Grieson Ronald, ed., Public and Urban Economics, Lexington, Mass. : Lexington Books, 1976
- TIEBOUT Charles, *A pure theory of local expenditures*, Journal of Political Economy, 64 : n° 5 : 414-424, 1956.
- McGUIRE Martin, *Private good clubs and public good clubs : economic models of group formation*, Swedish Journal of Economics, 74, n° 1, 84-99, 1972.

(35) The results discussed would not hold if the marginal costs of new network participants drops continuously more than their marginal benefit to an existing network user. The latter is unlikely since marginal cost, beyond a certain range, is either flat or very slowly decreasing, or in fact increasing.

(36) BUCHANAN, op. cit.

But beyond that point, the pro-expansion policy creates incentives to form alternative networks. And the more successful network policy is in terms of achieving universal service and "affordable rates", the greater the pressures for fracture of the network. Hence, the very success of network expansion bears the seed of its own demise. This is what I call the "tragedy of the common network", in the Greek drama sense of unavoidable doom, and borrowing from the title of J. Hardin's famous article "The tragedy of the commons" on the depletion of environmental resources. In the case of telecommunications the tragedy is that the breakdown of the common network not caused by the failure of the system but rather from its very success - the spread of service across society and the transformation of a convenience into a necessity.

■ Trends enabling network fracture

There are several broad trends that contribute to new network coalitions becoming an increasingly realistic proposition ; and they will be the subject of the next and more empirically-based section of this essay.

The saturation of basic service

The model argued that the expansion of a single network eventually leads to its fragmentation. What has been the reality of such expansion ? For a long time, the primary policy goal of most industrialized countries was to establish a network that would reach every household ; this also benefited the supplying industry. Even in highly developed countries the achievement of substantial network penetration (n --> N) is a very recent phenomenon. In Germany, penetration in 1960 was 12 %. In 1980, it was 75 % (37). In France, it was 6 % in 1967 (38) and 54 % in 1983 (39). Hence, the imposition of the costs of the last subscribers is a fairly new development, and responses to it are only now beginning to work themselves out. In the US, universal service was achieved some 25 years earlier.

Increasing cost of incremental subscribers

Over the years, lower-cost subscribers have tended to be added to the network earlier than high-cost subscribers. As the network reaches universality, connecting the last members increases cost. An indicator for rising costs : in the Bell System, the average capital investment cost per new telephone grew steadily (in 1982/3 dollars) (40).

---

(37) SCHULTE Josef, 1982, *Endgeraetekonzeption im fernsprechdienst der Deutschen Bundespost*, in Elias Dietrich editor, Telekommunikatión in der Bundesrepublik Deutschland, Heidelberg, R.v. Decker's Verlag, G. Schenck : 321.

(38) GUERARD A., LAFARGE G., PAUTRAT C., *Les régions dans la course au téléphone*, Economie et Statistique, n° 117, December, 37-49.

(39) AT&T Communications, 1983, *The world's telephones*, Morris Plains, NJ.

(40) Sources : Telecom Factbook, 1986, FCC, Statistics of communications common carriers, 1945, 55, 65, 75, 85.

1945 : $ 1928
1955 : $ 2050
1965 : $ 2580
1975 : $ 3960
1985 : $ 4624 (41)

An activist role by the equipment industry

Once universal penetration was reached, the industry had to reorient itself, because its activity level would have otherwise dropped dramatically.

Having been successful in spreading telephony, the supplying industry, too, became a victim of its own success. It was left with several strategies :

Option 1 : upgrade
After achievement of universal penetration, the equipment industry advocated an upgrade of the network. This means an accelerated supply push rather than demand pull.

It means move into videotex, ISDN, IBN, and cable television as ways to provide the industry with procurement contracts.

Option 2 : export
Increased attention to international activities can substitute for the shrinking basic domestic market. However, many of the more interesting markets in industrial and industrializing countries are protected by their governments who use the network as a way to promote a domestic electronic industry. The result are trade frictions around the world, and eventually partial opening of national markets in order to achieve reciprocity. This reduced the traditional territorial compartmentalization of the industry and loosened the close relation between equipment industry and network monopoly. One of the positive consequences is that prices fell even beyond the gains of technology.

Option 3 : targeting users as equipment buyers
Perhaps most importantly in the long run, manufacturers turned to the large users as a market for equipment. In the United States, whereas in 1975 virtually all of capital equipment in telecommunications was invested by the carriers, in 1986 it was only 2/3. About $15 billion were invested by non-carriers, mostly large users (42). Such equipment includes PBXs, multiplexers, concentrators, network management equipment, satellite and microwave facilities, etc.

Users increasingly assumed control over the network segments closest to them ; first, over equipment on their premises ; second, over the wiring segments in office and residential buildings ; it was natural, as the next step, that they began to share in a full array of telecommunications services within their building though "shared-tenant services", which shifted some of the switching from the public exchange to a common PBX and moved

(41) Encompasses all US carriers ; translated for new telephones from data on access lines using 1975 ratio.
(42) CRANDALL Robert W., *Fragmentation of the telephone network : implications for the policymaker*, Washington DC, Brookings Institution, 1988.

transmission from the public network to a private one. Next, users' each moved to switching and to local area networks (LANs) for high-volume links serving the data flows within an organization.

The implication is that the equipment industry, in the past a protector of the traditional monopoly, has increasingly become part of the process of creating alternative networks.

Reductions in equipment costs and increases in productivity

A fourth factor leading to new networks is that the underlying economics of transport and switching have shifted the cost curve for telecommunications considerably downward. A unit of communications has become much cheaper to transmit and switch. The cost of a network drops as electronic and photonic equipment becomes cheaper, more powerful, and lower in operating costs. Switch prices came down in price per line from $230 in 1983 to $144 in 1988 (43). Manpower requirements declined considerably (44), while productivity increased.

Similarly, the price per meter of fiber has come down from $7 in 1977 to 23 cents in 1988 while its transmission capacity has increased enormously ; LEDs dropped from $2000 a few years to $30 today. In 3 to 5 years fiber will be cheaper to install than copper.

For local distribution -in the past the segment with the greatest characteristics of "natural" monopoly- several different transmission technologies have emerged, including stationary digital radio, micro-wave, fiber-optics, and infra-red transmission.

In terms of the model, as the cost curve drops down as a whole, the critical mass point n_1, shifts to the left to a smaller minimal size, and it becomes easier to start an alternative arrangement.

Increases in user size

The communications volume of large users has gone up, and it takes fewer users to reach critical mass. Average use per line increases annually, on average, by about 4-7 %. For a large user, even a moderate percentage of expansion is likely to be a very large increase in absolute terms. Usage increases with the "informatisation" of society an economy. Information-based activities, including knowledge-intensive manufacturing and the growing white-collar service sector, contribute to the steady increase in telecommunications. This transition to information as a major activity is observable around the world in economically advanced societies.

The large users of telecommunications are corporate headquarters, banks, insurance firms, airlines, health delivery organizations, engineering and consulting firms, law offices, media organizations, and other providers of services. The shift towards such activity in highly developed countries was

(43) Nynex data, see in Noam Eli, *The past five years and the next five years : a look at the post-divestiture network and the next regulatory agenda*, Harry Trebing ed., Divestiture plus V and the coming regulatory agenda, forthcoming.
(44) New York PSC data see, ibid.

partly due to their loss of competitiveness in traditional mass-production vis-a-vis newly industrialized countries. It was also partly due to a large pool of educated people skilled in the handling of information. Information based services, including headquarters activities emerged as a major comparative advantage of developed countries. These activities were reinforced by productivity increases in information transactions through computers and advanced office equipment. In consequence, electronic information transmission, i.e., telecommunications, became of ever increasing importance to the new services sector. It also became a major expense. For Citicorp, America's largest bank holding company, telecommunications became the third largest cost, after salaries and real estate. This made the purchase of communications capability at advantageous prices more important than in the past, and in turn, led to the emergence of the new breed of private telecommunications managers whose function was to reduce costs for their firms, and who for the first time established sophisticated telecommunications expertise outside the postal-industrial coalition. These managers aggressively sought to establish low-cost transmission and customized equipment systems in the form of private networks of power and scope far beyond those of the past.

The increase in service employment is not surprising. Even if services were to remain a fixed share of consumption (in fact, though, they are increasing), their total would grow with GNP and their share in employment would rise, because their productivity increases more slowly than that of manufacturing and agriculture. Hence, special efforts are undertaken to increase service productivity, and this means a substantial role for information and communications.

In terms of the model, the growth of large users means that it takes a smaller number of them to reach any given network size n. This reduces transaction costs or organizing and coordinating a new network club, and makes it possible for a smaller number of users to enjoy the economies of scale.

Upward drift of cost curve of the old network

Costs and efficiencies of networks are not simply a question of engineering, but also of market structure. The existing network, operating as an exclusive arrangement, tends to drift upwards in terms of cost, for a given technology vintage. This can be exacerbated by regulatory arrangements that lead to wrong incentives, such as to overcapitalization (Averch-Johnson effect). Some indicators for this are the cost reductions achieved by US companies when competitive pressures started to make themselves felt.

For example, AT&T's equipment installation and maintenance cost was estimated at $61 per hour as opposed to $33 for IBM and $28 for MCI (New York Times, Aug. 5, 1984, p. 4F). For New York Tel, cost reductions per access line, following the AT&T divestiture, were considerable (45).

The implications are that a new network, unencumbered by the accumulated high-cost attributes of the old one, could operate on a lower cost curve even in the absence of technical progress.

(45) See NOAM, op. cit., *The past five years and the next five years...*, forthcoming.

Specialization

As the information flow requirements of large users become still larger, they are also becoming increasingly specialized. Equipment offered by numerous vendors permit many configurations to accommodate the requirements and procedures of organizations. It is no longer as necessary to forgo benefits of specialization in order to benefit from cost sharing.

By their very nature and tradition, the traditional networks provide standardized and nationwide solutions, carefully planned and methodically executed. In the old days, sharing a standardized solution was more acceptable to network members, because the consequential loss of choice was limited and outweighed by the benefits of the economics of scale gained. As the significance of telecommunications grew, the costs of non-optimal standardized solutions began to outweigh the benefits of economics of scale, providing the incentive for non-public solutions. Furthermore, some users aggressively employed a differentiation of telecommunications services as a business strategy to provide an advantage in their customers' eyes, and they affirmatively sought a customized rather than general communications solution. While these considerations are most important for large firms, the differentiation of communications needs for small business and residential users has been moving apace, too.

The desirability to opt out of the traditional sharing arrangement depends, among others, on traffic densities. Where these are high, private networks of different types emerge, depending on degree of traffic specialization.

For high density general purpose use, private networks emerge, while for low-density general purpose the shared public network may last longest. For intermediate density, "virtual private network" provide an intermediate option. For specialized services, the network type to emerge, depending on traffic densities, are strategic private networks, closed use group networks, and value added networks. These classifications (46) can be charted as follows :

Traffic density Specialized	Network specialization General	
Dense	Private networks	Strategic private networks
Intermediate	Virtual private networks	Closed use group networks
Non-dense	Public networks (ISDN)	Value-added networks

As traffic densities and usage specialization grow, the core of the traditional public network -the lower left corner- is shrinking.

(46) This classification extends those of ERGAS Henry (Joan, get cite for Ergas paper in Noam & Komatsusake, etc.).

The pluralist network

The new network coalitions

These incentives to group formation can lead, where they are permitted to do so legally, to alternative sharing arrangements of new "network associations". It is a process that we called "the tragedy of the common network", because it is not the failing of the traditional system, but rather its very success which undermines it. The success of a communal network creates over-expansion, cost-shifting, and the forces for particularism. The largest of users increase their electronic communications to huge volumes, and their technical requirements are increasingly differentiated from those of average users. Because the combined volume of large users has risen so much, they can account for much of cost savings of sharing just between themselves. They form alternative network associations for large parts of their communications needs, first in-house, then with their closest suppliers, customers, or market partners.

Examples are large private intraorganization networks, shared tenant services, local area networks, wide area networks, and other specialized services.

These groupings of users need not be territorial. The idea of telecommunictaions as consisting of interconnected national systems is likely to be transcended in many instances, and specialized transnational networks will emerge. This becomes possible with the drop in cost of international circuits.

For satellite transmission, in particular, the marginal cost with respect to distance is close to zero. Fiber-optic links have also lower distance-sensitive costs. The implications are that communication flows can be routed in indirect ways to exit previous shared arrangements, or in order to join new and more congenial ones. Arbitrage becomes easily possible. This undermines attempts to administratively set rules for prices and service conditions.

In the future it is likely that specialized global networks will emerge for a variety of groups that communicate with each other intensely. Their relation to each other is functional rather than territorial, and they can create global clustering of economically interrelated activities much in the way that in the past related activities clustered physically near each other.

This has begun with the private networks of large users. These private networks, some of whose operation and administration require hundreds of skilled technicians and managers, begin to carve out slices from the public network. It does not take a large number of private networks to have an impact. In the largest 3 % of users typically account for 50 % of all telephone revenues. These activities are spearheaded by private firms, but are not exclusive to them ; non-profit institutions such as hospitals and universities, and public organizations such as state and local governments are also actively pursuing similar strategies. Similarly, groups have begun to interlink through private VAN networks such as SWIFT.

Examples for group networking are : advertising agencies, media firms, printers, chemical manufacturers, environmental protection agencies, insurance agencies, hospitals, record rooms, police.

In some instances these will have special performance features that distinguish then from the general "public" network. In the first example of the list, network bandwidth probably must be quite high to permit transfer of high resolution graphics.

In other instances, additions of supporting software and data bases provide a more powerful communication, as in the second example. But in many instances, such as the third, it is probably the price of inter-communications that drives the arrangement.

Many entities are likely to participate in several networks. Furthermore, the pluralist network does not mean separate transmission links for each subnetwork at every point. It will often make sense to transport the traffic of several low volume users part of the way on the general network until the point where there is enough aggregate traffic to branch off. The economics of sharing are not abolished. But they must prove to be superior as a matter of choice rather than being imposed by a legal requirement.

Interconnection and Open Network Architecture

As the earlier discussion of interconnection of old and new networks demonstrates, it is not necessary to cut one's links to other participants. Multiple networks need not lead to multiple telephone sets on one's desk. This is not more necessary than having multiple currencies if there are several banks in a country. In the situation of multiple networks, the functioning of the network federation depends on rules of interconnection. Such rules are being structured in regulatory approaches of Open Network Architecture (ONA) in the United States and of Open Network Provision (ONP) in Europe. Such arrangements provide additional impetus for new networks in that they permit a the gradual establishment of a new network. They also permit a wide range between a full an separate physical network, and the use of an existing network in different and innovative ways, somewhat apart from the general "public" usership.

The concept of gradualism inherent in Open Network Architecture is important ; many advocates of the traditional shared network system believe that the demands of pluralism could be met by a flexibility for software network options, without altering the exclusivity of the physical network of the traditional monopolist. This is wishful thinking. Granted, to permit software networks on a transmission monopoly is a correct first response to the emerging pressures. However, it is unlikely to be adequate in the long run. Soon, users will want to supplement transmission offerings with additions to satisfy their preferences in terms of technology, control, and economics. The exclusive network cannot be the superior solution in each instance, particularly if it has to follow political mandates, or if it cannot bargain individually on prices.

It may be asked why the public network provider could not supply each user association with whatever it needs, without requiring new physical network arrangements. Theoretically this is possible, and indeed some of the change is taking place on private networks supplied by the monopolists. But institutionally it is unlikely to be adequate, because it would, first, require heroic willingness by the traditional network to collaborate with schemes designed to reduce its revenue. Second, it would require a substantial lowering of its cost-structure which has crept up over time as suppliers and employees shared in the monopoly profits. And third, it would require an enormous upgrading of innovativeness and speed, and the traditional large organizations are often not best equipped for that.

■ International instability

This then leads to the question whether the process of network evolution in the network is expansionary internationally in the sense that one country's developments affect the others.

One could of course argue that somehow the strength of one country is the cause of change elsewhere ; in this instance, the United States would be the country affecting the others. But this would require an explanation why the US private model did not affect Western Europe and the Pacific in the first decades after World War II, when the US was in relative terms much stronger than today. More likely, there are common elements that lead to change in each country domestically. But there are also factors of interaction. No country can be truly an island anymore.

The more interrelated countries and economic activities are, the less likely are there stable solutions to separate policies. And where instabilities exist in one country, they affect the entire system. It becomes increasingly difficult to control all of the elements in such a complex matrix of interrelations. Ultimately to maintain the traditional arrangements, it would be necessary to give control to a supra-national body over many countries and many economic activities. But such power would be very untraditional, and this is the dilemma of traditional policy. Since such supra-national power does not exist, is usually not deemed desirable, and is hard to enforce, regulatory strictness unravels.

Hence, in an interrelated world, the politically optimal regulation may be different than for a single activity in an isolated jurisdiction.

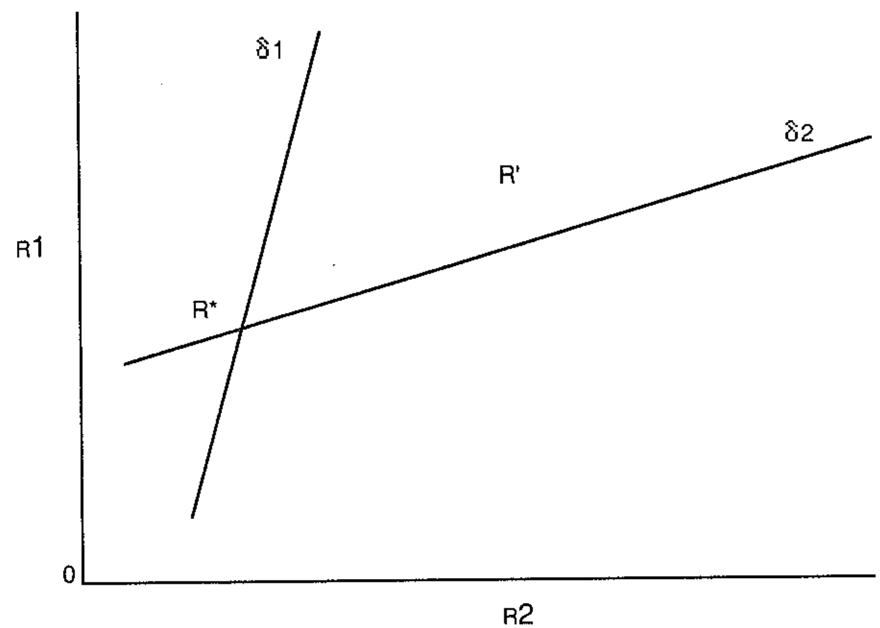
In most instances, one would encounter a positive cross elasticity of regulation, that is, liberalization in A leads to greater liberalization in B, and vice-versa. But in some instances, the cross-elasticity would be negative. If one country's monopoly exercises restrictive policies, its firms will be disadvantaged internationally, and foreign firms may choose not to domicile themselves in the country. Similarly, acquaintance with options available elsewhere create, by their demonstration effect, pressures for change across borders.

An example are transborder data flow protection laws. The less protected data is in one country, the tighter the other may become in response. And this is indeed what has happened in Europe. This can lead to instable equilibria.

We can think of two "reaction functions" f_1 and f_2 (in Graph 3) which track the response of one regulation to the other's given level.

A point such as R' would denote the two independently set regulatory policies. But once we postulate reactions to each other, there would be a gradual shift that converges in R*.

Graph 3



Note that there is no need for coordination between 1 and 2 ; an equilibrium can be reached by unilateral actions and reactions. However, for an equilibrium to exist requires that the reaction functions 1 and 2 are such that 2 is steeper than the inverse of 1 at the point of intersection, i.e. that δ_2, cuts δ_1 from below.

If the reverse is true, there is no equilibrium and the regulatory strictness either moves successively higher or lower to corner solutions. This means excessive deregulation or regulation due to international interaction. For other configurations of the reaction functions, cyclical change is possible.

Instability raises questions of how to prevent it, and therefore leads to the issue of policy coordination. But even in stable situations, R* may not be the optimal result for either 1 or 2, or for both, and policies may be sought to affect moving to R* back to R', such as international cartel agreements.

For a country to participate in such supra-regulation, the advantage of the limination of uncontrolled interaction must be greater than the value attributed to control and independence. Of course, de facto independence already had been lost through the mechanism of interaction, and supra-regulation reflects this.

The problem of any coordinated regulatory structure is its instability. First, there may be no equilibrium possible, because the reaction functions do not meet the stability criterion described above. Second, one jurisdiction's adherence to an agreement provides the other with an opportunity of gain by seeking a non-cooperative policy. In each jurisdiction there are pressures to seek one's own ideal regulatory level, which is likely to be different from the agreed upon level or from the interactive equilibrium. Going-it-alone can be due to short-sightedness or lack of understanding of the interaction involved. But it can be based on the rational desire to gain advantages over others by breaking joint policy, at least in the short run.

In telecommunications, communications "havens" may emerge. The example of telex service is instructive. In the 1980s, London-based telex-bureaus started to retransmit traffic between North America and Continental Europe in defiance of CCITT cartel "recommendations" against such retransmission. It was profitable for UK firms to break these rules, since this generated more traffic and made the UK more attractive as a business location. In time, the cartel rules were held to be illegal.

It is thus important to recognize that domestic instability is linked to international instability in regulation. That is, if a country's domestic system unravels, as has been discussed, it leads to repercussions of adjustment elsewhere, which in turn affect the other countries' domestic stable solution.

These changes lead to unstable situations that affect the entire system. Network diversification can have multiple secondary effects, which in turn can lead to further diversification. At the same time, collaborative regulatory adjustments becomes more difficult, because they cannot be confined to sub-sectors.

Applied to telecommunications, one should therefore expect the trend towards sub-networking to be spreading, though accompanied by efforts to stabilize its collaborative aspects. But this is not an entirely one-sided process. The United States has been far less liberalizing in international services than in domestic ones, for example in the licensing of rival international satellite carriers, due in part to the resistance of other countries. Hence, greater internationalization of communications will make it more difficult for the US to blithely go its own way. As the matrix of interrelations becomes steadily more cross-elastic, one can also have some oscillations. But the overall tendency, in the long term, should lead to reduced protection of the traditional network system internationally. In that sense, network pluralism is an expansionary process. It is not so much an ideological choice, but a response to an internal inability to structure a stable equilibrium that serves the multiple domestic interests and goals, and one has to predict that similar instabilities will spread throughout the system.

In the past, international interactions have often been used to stabilize domestic arrangements. Monopoly telephone administrations often used international agreements of ITU and CCITT to shore up occasional domestic dissatisfaction. But the symmetrical scenario is being played out now, with the opposite direction, as international trends of liberalization undermines domestic stability.

■ Outlook

The model of this essay identified an evolution of network development with three stages :

The cost-sharing network
Expansion of the network is based the logic of spreading fixed costs across a large number of participants, beyond a take-off point ("critical mass"), and of increasing the value of network connectivity.

The redistributory network
At this stage, the network grows beyond the size that is optimal to its original members through politically directed expansion caused by the formation of internal coalitions that lead to transfers from some users to others.

The pluralistic network
Beyond a certain point, the cohesion of the unitary network breaks apart because the dynamics of expansion and redistribution leads to a divergence in the interests of its participants that cannot be reconciled anymore within one network. The results are exit, formation of new networks, and the emergence of a federation of subnetworks. The network has progressed to its "tipping point", where its cohesion breaks up and a multi-networks system emerges.

These trends have a certain logical progression. At first the network expands because it makes economic and technical sense. Later, because it makes political sense. But as the network provider succeeds in providing full service to every household, it also undermines the foundation of its exclusivity.

Most countries in the world are still in the phases of the cost-sharing or redistributory networks. A few have reached substantial universal penetration of telecommunications for a number of years, and it is here that the transformation of the network system has begun progressing to its third stage. Since there is a strong correlation of economic growth and telephone penetration -roughly an additional telephone per $50,000 GNP- countries with high economic growth are likely to progress rapidly through the first two stages towards the third. Eventually, through economic growth and through the instabilities transmitted from the more advanced countries, many other countries will move into a pluralistic network environment, and be faced with the policy issues inherent to such a system.

Where does all this lead to ?

It leads, first, to diversification and disintegration. But in a broader sense, it leads to normalization. Normalization means that telecommunications network provision will resemble much of the rest of the economy : no monopoly, some redistribution, some international provision, a mix of public and private providers, and some regulation.

The network environment is likely to become essentially a pluralistic network of user associations, or network of networks which are part overlapping, part general, part specialized along various dimensions such as geography, price, size, performance, software value-added, ownership status, access rights, specialization, etc. This is not to say that economies of scale and scope will become irrelevant : there will still be broad-based public networks, and powerfully integrated networks with broadband capability. But just as important will be the economies of group specialization and of clustering. These differentiations will permit users with similar needs, or with frequent interaction, to operate on more efficient networks. it will also permit the traditional networks to be more efficient for their members, since they need not be all things to al people. On the other hand, there may well be market failures inherent to a pluralistic network environment that will retard or prevent investments in new focus of networks.

Where does such segmentation leave future policy makers ? It would be naive to expect less regulatory tasks. Many disputes become less intraorganizational within the single network and more public in nature. The main regulatory tasks which the pluralistic network system raises for the future would deal especially with "inter" issues : inter-operability, interconnection, integration, inter-media, international. These issues include :

1. Protection of technical compatibility.
2. Protection of physical interconnection and service access.
3. Protection of free-flow in information content such as embodied in common carriage principles.
4. Establishment of new mechanisms of inter-network redistribution as a substitute for the intra-network cost-shifting, in order to assure the viability of a backbone network.
5. International coordination of national policy arrangements to match the global scope of networks.
6. The prevention of oligopolistic behavior and of cyclical instability.
7. Protection of internetwork service quality and privacy.
8. Consumer protection in a system of multiple and private networks.
9. Assurance of optimal network investment in an environment where market failure may exist.

None of these tasks is beyond our grasp in terms of complexity or political feasibility. But they require us to end the nostalgia for the simplicity of the golden age, and to imagine a very different network environment.

Externalités de blocage et processus de diffusion : le cas de la téléphonie mobile (*)

Frantz Rowe, Etienne Carpentier
Département Economie et Management
Ecole Nationale Supérieure des Télécommunications

Le développement économique actuel se caractérise à la fois par une différenciation toujours plus grande des produits et services et par l'incorporation toujours plus rapide de nouvelles techniques. La différenciation accélérée des produits résulte de l'avivement de la concurrence et de l'élargissement des marchés auxquels répondent les politiques marketing, tandis que le progrès technique, principalement fondé sur ceux des technologies de l'information, raccourcit encore davantage le cycle de vie des produits.

Le cas de la téléphonie portable paraît caricaturer cette situation puisqu'aujourd'hui les industriels et les opérateurs des télécommunications connaissent déjà les trois générations de technologie numérique qui remplaceront celle du téléphone sans fil analogique et étendront considérablement ses possibilités d'utilisation. Après en avoir brièvement rappelé les caractéristiques, l'objet de cet article est d'analyser les interdépendances et les enchaînements possibles de ces technologies afin de proposer des scénarios du processus de diffusion des mobiles. En particulier, on insistera sur le rôle des externalités négatives, qualifiées ici d'externalités de blocage, qui se manifestent dans les sytèmes de communications avec les mobiles et sur leur répercussion dans les trajectoires technologiques possibles suivies par les opérateurs.

Caractéristiques techniques

Généralités

Pour aborder la téléphonie mobile, certaines notions techniques doivent être exposées. La première d'entre elles est la notion de cellule. Le territoire couvert par le système technique est découpé en portions appelées cellules, au centre desquelles se trouve une antenne réceptrice/émettrice. C'est avec cette antenne que les mobiles, placés dans la cellule, se mettent en communication.

La couverture continue d'un territoire donné requiert la juxtaposition, voire le recouvrement, des cellules suivant le schéma théorique hexagonal ; en

(*) Cet article résulte d'une recherche menée en collaboration avec les départements Electronique, communications et réseaux de l'ENST que nous remercions. Cette recherche a été financée par le service de Télécommunications avec les mobiles de France Télécom et a largement bénéficié des conseils de M. Brussol. Nous remercions enfin L. Benzoni pour ses remarques constructives.