500

Why Networks Change:
A Theory of Network Evolution

Eli M. Noam

Columbia Institute for Tele-Information
Graduate School of Business
809 Uris Hall
Columbia University
New York, New York 10027
(212) 854-4222

Why Networks Change:

A Theory of Network Evolution

Eli M. Noam

Columbia University

Preliminary

I. *INTRODUCTION*[1]

Recent years, have witnessed major transformation in telecommunications industry structure, from monopoly to an increasing diversified environment. Because many of the changes originated in the United States, they are often viewed as the product of particularly American business interests and ideology. But more recently, several other industrialized countries have begun to adopt similar policies, or at least to discuss changes that previously seemed unthinkable.[2]

These developments raise the question whether the change has explanations that are 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.

---

[1] I am grateful to Bruce Greenwald and Julianne Nelson for their helpful comments.

[2] The United Kingdom, for example, has created a telecommunications duopoly and allowed Mercury to compete with a privatized British Telecom for local and long-distance service; Germany has removed its telecommunications operator, Deutsche Bundespost, from the Ministry of Posts and Telegraphs, leaving the latter a regulatory role only; Japan privatized its telecommunications monopoly service provider, NTT, and has allowed entry into all telecommunications sectors, including local service, under various conditions; and the European Community has called for the separation of regulatory and operations functions for all member states, and has mandated a liberalized entry policy for value-added services. See generally, Eli M. Noam, *Telecommunications in Europe*, New York: Oxford University Press, 1991.

Unfortunately, there has been little attempt at a broader-gauged interpretation of the formation and transformation of networks that can explain the dynamics of change.[3] To provide such analysis is the aim of this essay.

II. *Theories for the Emergence of Multiple Networks*

A number of explanations have been offered for the demise of monopoly in telecommunications. There are four major types of theories with different drivers: technology, politics, non-sustainability, and market structure.

The *technological* explanations stress new transmission options that lowers entry barriers, and the merging of telecommunications and computing which undermines monopoly power.[4] But these observations are not adequate explanations. The same technologies are

---

[3] One attempt was a US Department of Justice report on the post-divestiture network. (Huber, Peter, The Geodesic Network. Washington D.C.: U.S. Government Printing Office, 1987.) Another approach is that of Koichiro Hayashi of NTT. (The Economies of Networking - Implications for Telecommunications Liberalization, paper presented at the IIC Conference, Washington, D.C., Sept. 1988).

[4] See, for example, Ithiel de Sola Pool, *Technologies of Freedom*, Cambridge Mass.: Harvard University Press, 1983 ("This development [diminishing importance of distance to cost], along with the development of multiple technologies of communication and of cheap microprocessors, will foster a trend toward pluralistic and competitive communication systems." p. 229); George Gilder, *Microcosm: The Quantum Revolution in Economics and Technology*, New York: Simon & Schuster, 1989; J.S. Mayo, "The Evolution of Information Technologies," in *Information Technology and social transformation*, B.R. Guile, ed., Washington, DC: National Acadamy Press, 1985; S.F. Starr, "New communications technologies and civic

(continued...)

available anywhere on the globe, and certainly in the developed world, yet their impact on network structure has been highly varied, providing no evidence for a technological determinism at work. Technology offers the precondition for necessary but not sufficient institutional change.

Political explanations use the perspective of countervailing powers, arguing that in the information age, a telecommunications monopoly becomes too powerful so that its scope needs to be limited by a governmental structural policy to establish competition.[5] The problem with this view is that the creation of a multiplicity of carriers is not the only policy option. Alternatives might well be a stricter nationalization, or more effective regulation, or a size-reduction along geographical and/or functional lines while maintaining monopoly. Thus, it is not clear why the introduction of competition should be the result of monopoly power.

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

---

[4](...continued)
culture in the USSR," Paper presented at the Center for International Affairs, Harvard University, October 1987 (arguing technological options and convergence of data processing and transmission made pluralization inevitable even in centrally planned economies).

[5] EN: Having searched high and low, I have not found any cites for this position. Do you have a name or two in mind?

regulation. This view is essentially that of an economic *non-sustainability* theory.[6] This 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 restrictive assumptions.[7]

Another type of explanation is the classical industrial organizational view. It postulates that monopoly structure leads to inefficiency in performance, and hence eventually to the entry of competition. Yet this view 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 U.S. 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 responsibility.

---

[6] Baumol, William J., Panzar, John C., and Willig, Robert D., Contestable Markets and the Theory of Industry Structure, (New York: Harcourt Brace Jovanovich), 1982.

[7] 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).

Thus, none of these theories for the emergence of multiple networks provides an adequate explanation.

In contrast, this paper advances an alternative view based on the dynamics of group formation. Its explanation is not based on the *failure* of the existing monopoly system. To the contrary, the breakdown of monopoly is due to its very *success* in advancing telephone service and in making it universal and essential. But as the network expands, political group dynamics take place, leading to overexpansion, redistribution, and instability. This creates increasing incentives to exit from the "sharing coalition" of the network, and to an eventual 'tipping' of the network from a stable single coalition to a system of separate sub-coalitions.

This view of the effectiveness of monopoly, yet of its 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.

III. *A Model of Networks*

1. The basic model[8]

One can look at a network is as a *cost sharing arrangement* between several users. Let the total cost of a network serving n subscribers be given by C(n).

n assumes that users are homogenous. (Of course, some network participants are much larger than others, but that poses no problem if one defines a large organization to consist of multiple members of type n, e.g., telephone lines or terminals rather than subscriptions. Later, we will drop that 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. We assume network externalities to exist, ($\delta u/\delta n > 0$), though at a declining rate, i.e. a subscriber is better off the more other members there

---

[8] I will follow the analysis in Noam, Eli, "The Next Stage in Telecommunications Evolution: The Pluralistic Network," paper presented at the Pacific Telecommunications Conference, Japan, October 1988. For sections 1-4, I adapt part of the methodology of my colleague Geoffrey Heal, "The Economics of Networks," Columbia University, unpublished paper, 1989, for which I am indebted.

The model of the present paper can also be used for "standards coalitions" rather than "network coalitions." For the literature on standards, see David, Paul A., "Some new standards for the economists of standardization in the information age," in P. Dasgupta and P.L. Stoneman, eds., Economic policy and technological performance, Cambridge University Press, 1987b.

are on the network, *ceteris paribus* (including network performance and price).[9] Price is also a function of network size, since cost is shared by network uses. For simplicity, utility is expressed in monetary units

$$u = u(P) + u(n) = -P(n) + u(n) \qquad (1)$$

We assume that the network membership is priced at average cost, i.e. that users share costs equally.[10] (This assumption will be dropped later.) This can be shown

---

[9] 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. Network externalities are discussed in Brown, Stephen and David Sibley, The Theory of Public Utility Pricing, Cambridge University Press, 1986. See also W.W. Sharkey, The Theory of Natural Monopoly, Cambridge, U.K.: Cambridge University Press, 1982.

[10] More completely, if a user's budget y goes toward network access P and other goods x (with $P_x = 1$)

$$y = P + x \qquad (2)$$

Then $u = u(n, y-P)$

With average cost pricing, thus becomes the indirect utility function

$$u = \underset{n}{u}(n, y - \underline{C(n)}) = \underset{n}{u}(n) + y - \underline{c(n)} \qquad (2)$$

if utility is expressed in monetary units.

schematically in Graph 1, where u(n) is steadily increasing, though at a declining rate, and P = AC = C(n)/n is declining, at least at first.

2. Critical mass

Subscribers might find it attractive to join a well-sized network to share costs, while the number of subscribers n adds to utility. This can be seen in Figure 1, where the 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_1, a network will not be feasible, unless supported by external sources. We define critical mass as the smallest number of users such that a user is as well off as a non-user $u(n) = P(n)$.[11]

To reach n_1 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_1 will be reached within the range $n < N$, where N = total population, and within the range of demand. 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] Heal defines it, similarly, as

$$u\,(y/p - [F + f(n)]/np;\, n) = u(y/p;\, 0).$$

See, also, David Allen, "Net telecommunications service: Network externalities and critical mass," *Telecommunications Policy*, September, 1988, pp. 257-271.

3. Private Optimum

Through the cost-sharing phases of network growth, the earlier network users can lower their cost by adding members. However, average cost increases in the range beyond the point n_0 where AC = f'(n).

Beyond n_0 expansion becomes unattractive for cost reasons; new subscribers, for example because they are in more remote locations with lesser population density, are more costly to serve.[12] However, some further expansion would be accepted by the network members since newcomers beyond the low cost point would still add to utility. The optimal point n_2 is given where the equation holds

$$u'(n) = \frac{1}{n}[C'(n) - C(n)/n] \qquad (4)$$

This is the case in the range of increasing AC ((C'(n) > AC), since u'(n) is positive). If they are given the ability to exclude, existing subscribers would not accept network members beyond n_2, the private optimum.[13]

---

[12] E.g. [data]

[13] This is not to suggest that such self-restriction in size actually exists. Almost always is there a governmental requirement for expansion instead of genuine self-government of users. But an example is Bolivia, where local subscribers are members of cooperatives, and have resisted an expansion that reduces the value of the their membership shares.

4. Social optimum

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

Assume social welfare given by the sums of utilities[14]

$$W = n[u(P(n)) + u(n)] = n[-C(n)/n + u(n)] \tag{5}$$

so that its derivative

$$\frac{dW}{dn} = -C'(n) + n\,u'(n) + u(n) = 0 \tag{6}$$

$$u'(n) = \frac{1}{n}[C'(n) - u(n)] \tag{7}$$

Marginal cost equals the incremental user's utility, plus the existing n user's marginal utility. Since $u(n) > C(n)/n$ below the a point of intersection n_4 in Figure 1, social optimum n_3 is greater than private optimum n_2. (It should be noted that the same size will be chosen by an unconstrained monopolist that sets the price at $P = u(n_3)$ to exhaust consumer surplus.)

---

[14] We assume that the utility off network is equal to income y of the budget constraint.

$$u(o,x) = y$$

5. Entitlement Point and Universal Service Obligation

The discrepancy of private and social optimum leads to government intervention, normally known as a "universal service" policy.

To understand the politics of government-directed network expansion, let us assume a political decision mechanism in which the majority rules the single network. As a first case, assume that private optimum size $n_2 < N/2$, which means that 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, especially in the allocation of the fixed cost. It will be shown in sections 8 and 9 that this coalition formation will lead to an over-expansion of the network.

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 benefits are negative, leading to attempts at exclusion. In economic terms, the

argument is made to expand the network at least to where C'(n) = P = u(n), leaving fixed costs to be distributed unequally, for example, by a Ramsey pricing rule. When the marginal private 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 thus define n_2 as the "entitlement point."

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

6. Exit From the Network

There may well be a point where the network is expanded by government requirement to an extent that, given its increasing cost, a user is better off by not participating. We define n_4

as the "exit point," i.e., the largest n such that an indifference exists between dropping off the network and sharing in the cost of supporting the expanded network.

$$u(n) = u(P). \tag{8}$$

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. Thus, assuming $n_4 < N$, a government's aim to establish a truly universal service is normally infeasible without resorting to a subsidy mechanism or price discrimination. In other words, a universal service policy is dependent on a redistributive policy.

7. Political Price Setting and Redistribution

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

As has been shown above, 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.[15] Let us assume at 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 by the 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. It is given by the utility u(n) and a factor k(n) to account for users' inability to adjust to a sudden absence of service in the short and middle term. K account for the assymmetry in entry and exit. Once one becomes a member of a network, the desirability of leaving it is larger than the utility of joining had been originally and if price gets pushed above u(n) minority + k(n),

---

[15] This analysis should not suggest that a self-governing and voting mechanism exists 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.

subscribers would drop off. The majority bears the rest of the cost. The minority's price P_B will be such that $P_B = u(n_2) + k(n_2)$. The majority's price will then be[16]

$$P_A = (2/n_2)C(n_2) - u(n_2) - k(n_2) \quad (9)$$

This then is the redistributory outcome, assuming no discrimination within majority and minority, and a fixed network size n_2.

8. Monopoly and Expansion

But such redistribution and size are not a stable equilibrium, for several reasons. As prices to the minority are pushed up to the limit and beyond, there are now incentives for the minority network members to exit the network and form new ones in which they would not bear the redistributory burden. This exit would deprive the majority of the source of its subsidy and is therefore undesirable to it. The way for the majority to prevent this "cream-skimming" or "cherry-picking" is to prohibit the establishment of another network. There are also incentives for arbitrage from low price to high price users. This, too, requires prohibition and enforcement. 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 is unstable insofar as it will expand beyond n_2. For the majority, there is added marginal utility from

---

[16] We use in the following the continuous n/2 rather than the discrete (n/2) - 1 and (n/2) +1.

added network members, while much of its cost is borne by the minority. The majority will therefore seek expansion. Initially, the majority would admit new members up to the point n_5 where marginal utility to its members is equal to the marginal price to them, subject to the maximum price extractable from the minority. 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 ($n_5/2$) 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 a majority member maximizes welfare, W_A.

$$\frac{du}{dn} = o, \text{ where } W_A = u(n) - P_A \qquad (12)$$

substituting from (9), we have

$$\frac{du}{dn} = u(n) - (2/n)C(n) + u(n) + k(n) \qquad (13)$$

$$\frac{dw}{dn} = 2u'(n)-(2nC'(n)-C(n)/n^2) + k'(n) = 0 \qquad (14)$$

This expression is positive at the private optimum n_2, leading to an optimum size

$$n5 > n2 \qquad (15)$$

The difference in size varies with k. The greater the cost of dropping off the only network, the larger the network will become through redistribution. Protective rules of monopoly create a high k, and so does the greater dependency on network participation. However, if in the process $n_5/2$ becomes larger than the critical mass point n_1

(defined by $u(n1) = \frac{c(n1)}{n1}$)

they could drop off and create a new network.

9. Network Tipping[17]

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. 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.

---

[17] The terminology of "tipping" is due to Schelling, Thomas, *Micromotives and Macrobehavior*, (New York: W.W. Norton, 1978).

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_5 would not be optimal to the majority anymore, and members would have to be forced out. And this, in turn, would 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 the minority is too small to create a self-supporting new network. One might call this the effect of *potential exit* by the minority, and it results in a lessened redistribution.

10. Unequal User Size

We have assumed so far that 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 $n = Av^{-a}$ (16)

where $A > 0$, $a \geq 1$

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

$$n_m = \int_{n/2}^{n} A\, v^{-a}\, dv = (A/(1-a))\,(1 - n/2^{1-a}) \qquad (17)$$

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.

11. 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 $n_A + n_B$, while not being subject to redistributory burden. This is why interconnectivity is a critical issue for the establishment of alternative networks, as the historical examples demonstrate, from the *Kingsbury*

Commitment in 1913[18] to *Execunet* in 1977[19] and today's *ONA*[20] and New York's *collocation*[21] proceedings.

Since the benefits of network reach remain minority subscribers', exit decision becomes strictly price-driven, and takes place if utility plays no role, price reach remains the same through interconnection.

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 its users.

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

---

[18] Letter of Nathan C. Kingsbury, vice-president of AT&T, to James McReynolds, U.S. Attorney General, dated December 19, 1913 (AT&T agreed to connect independent telephone companies to the AT&T network, among other provisions, as a compromise to avoid antitrust litigation).

[19] *MCI Telecommunications Corp. v. FCC*, 561 F.2d 365 (D.C. Cir. 1977) (*Execunet I*); see also, *MCI Telecommunications Corp. v. FCC*, 580 F.2d 590 (D.C. Cir.) (*Execunet II*), *cert. denied*, 439 U.S. 980 (1978).

[20] *Third Computer Inquiry*, 104 FCC 2d 958 (1986), *modified*, 2 FCC Rcd 3035 (1987), *further reconsid. denied*, 3 FCC Rad 1135, *vacated and remanded, California v. FCC*, 905 F.2d 1217 (9th Cir. 1990).

[21] Opinion No. 89-12, *Opinion and Order Concerning Regulatory Response to Competition*, Case 29469, issued May 16, 1989, at 24-29.

mutual interconnection charges. Interconnection facilitates the emergence of new networks. It lowers entry barriers. On the other hand, it may reduce competition by establishing cooperative linkages instead of end-to-end rivalry.[22] Interconnection is a useful concept, because it responds to the often-made claim that a single network is necessary for universal reach. This is clearly incorrect. Interaction does not usually require institutional integrations, and this was one of Adam Smith's major insights. Otherwise, we would have only one large bank for all financial transactions.But as the next section will show, it also may lead to market failure in the establishment of the original network.

12. 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

---

[22] Mueller, Milton, "Interconnection Policy and Network Economics," paper presented at Telecommunications Policy Research Conference, Airlie, Virginia, Oct. 31, 1988.

"infrastructure," a concept centered around externalities. The question now 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_i, 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.

How could one offset this tendency if it is deemed undesirable? Patents are one solution. Where a service is innovative but not patentable, one might create a "regulatory patent" for a limited period of protection. 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 quite 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 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.

13. 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 be lower with the ensuing competition.

The welfare implications of the formation of collective consumption and production arrangements is something analyzed by theorists of clubs.[23] 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

---

[23] Schelling, Thomas C., Models of Segregation, Santa Monica: Rand, 1969. Buchanan, James M, and Tullock, Gordon, The Calculus of Consent, Ann Arbor, Mich.: The University of Michigan Press, 1965 [CHECK IF PROPER CITE]. Tullock, Gordon, "Public Decisions as Public Goods," Journal of Political Economy, no. 179: no. 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: no. 5: 414-424, 1956. McGuire, Martin, "Private Good Clubs and Public Good Clubs: Economic Models of Group Formation," Swedish Journal of Economics, 74: no.1: 84-99, 1972.

decision-making. The economically optimal association size need not encompass the entire population.[24]

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.[25]

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 re-distribute 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

---

[24] 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.

[25] Buchanan, op. cit., p. 4,5.

groups.[26] 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 or another system of revenue shifting such as access and interconnection charges, is more efficient than monopoly and internal redistribution.

14. Conclusion

The 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.

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 we can call the "tragedy of the common network," in the Greek drama sense of unavoidable doom, and borrowing from the title of G. Hardin's classic article "The Tragedy of the Commons"[27] on

---

[26] McGuire, XYZ

[27] Hardin, Garrett, "The Tragedy of the Commons," Science, vol. 162, Dec. 13, 1968.

the depletion of environmental resources.[28] In the case of telecommunications the tragedy is that the breakdown of the common network is 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.

File: AEA

Disk: BeefA

Draft Date: August 23, 1991

[28] Tragedy is used in the sense of Alfred North Whitehead: "The essence of traumatic tragedy is not unhappiness. It resides in the solemnity of the remorseless working of things."

Quantifying Private Networking: Definition and Measurement Problems

Milton Mueller

Columbia Institute for Tele-Information
Graduate School of Business
809 Uris Hall
Columbia University
New York, New York 10027
(212) 854-4222

QUANTIFYING PRIVATE NETWORKING: DEFINITION AND MEASUREMENT PROBLEMS.

Milton Mueller
International Center for Telecommunications Management
University of Nebraska at Omaha.

ICTM recently conducted a survey of 30 major telecommunications users in a mid-sized U.S. city. Among other things, the interviews collected quantitative data about the types of services and facilities used by the respondents, including private facilities. Having directly contacted only a sample of the city's telecommunications users, the survey's picture of the extent of private networking is not necessarily representative of that city, let alone the whole country. This survey alone consumed about two months of full time work. I leave it to the reader to calculate how long it would take to apply this method to the entire U.S.

I mention this experience for two reasons: first, it gave us a fairly good idea of what kind of definition and measurement problems exist; second, it gave us some real data to work with in this workshop.

Accurate quantification of private networking in the U.S. is a task of enormous proportions. This paper will be a discussion of <u>how to go about</u> quantifying the present extent of private networking in the U.S., not the actual quantification. The paper does, however, attempt to give

some empirically-based examples of the consequences of classifying things in different ways.

The purpose of this paper is to establish a conceptual distinction between public and private networks that is grounded in theory but is also capable of providing an operational basis for classification and measurement.

1. THE PUBLIC/PRIVATE DISTINCTION.

The CITI proposal ("Private Networking and Public Objectives") identifies in general terms a crucial and interesting research topic. But the distinction between public and private networks contained therein is not precise enough to sustain a quantitative analysis. Within its concept of "private networking," I find three different definitions used interchangeably, each with very different implications. These are:

D1. Ownership
D2. Access
D3. Sharing

D1. Ownership. Ownership of course refers to who holds legal title to the facilities. According to this definition, private networks are telecommunications systems owned by someone other than public carriers: large corporations, universities, state and federal governments, etc. To the extent that this principle sets up a distinction between The Public Network and

Everything Else, its sharpness is dulled by the fact that The Public Network is no longer a singular entity but (to use the proper Noam-enclature) a "federation" of many different local and long distance carriers, all privately owned. Still, ownership is a valid criterion in that few would quarrel with the statement that an organization which owns its own communications network can be said to have a private network.

Important as it is, ownership cannot by itself be the boundary line. Many large-scale private networks rely extensively on dedicated facilities leased from public carriers. Actual ownership, where it exists, is usually confined to private premises. The most common pattern is for large users to own the facilities located within their building or campus but to lease from carriers anything that crosses public rights of way. One very large user in our survey is having an RBOC construct a fiber ring around the city that will stop at all IXC POPs. The long-term contract gives the private company sole and exclusive use of the facilities, but legal title is retained by the RBOC. The federal government, for that matter, does not "own" FTS-2000 facilities, and yet it is, by any stretch of the imagination, a private network.

D2. Access. "Access" refers to whether the user group connected by a network is "open" or "closed." Once again this is a distinction that makes sense only in comparison to the universal public switched network. An open network means that there are no predetermined limits on who can and cannot join the network. Reciprocal access is achieved by everyone who applies and pays the subscription charge. A closed network restricts access according to discrimination criteria other than the simple uniform "entry fee" charged by the PSN. It may be restricted to affiliates of a corporation, an industry group, etc.

Restricted access is an important but not defining characteristic of private networks. Private networks do not necessarily involve closed user groups. The internal telephone system of the University of Nebraska at Omaha, for example, is provided over its own PBX. While owned and managed by the university, the system is fully open to traffic to and from the public switched network. True, ownership gives the university the power to establish conditions blocking incoming and outgoing traffic. But is it reasonable to argue that the university's choice whether or not to exercise this power alters its status as a "private" or "public" system? Facilities may be wholly owned and operated by a private user but be entirely open. By the same

token, the presence of closure does not necessarily mean that a network is private. Software-defined networks can restrict access, but such service can be offered by a carrier on a public, tariffed basis, and may not even involve dedicated facilities.

Inasmuch as the phenomenon of closed user groups represents the withdrawal of specific "communities of interest" (see Rohlfs, 1974) from the Public Switched Network, the issue is not really the "privateness" of ownership or facilities but the fragmentation or reordering of the network coalition (see Noam, 1990). It would be perfectly legitimate (and quite interesting, in fact) to base the public/private distinction on the open/closed distinction alone. This would lead in a very different direction, however, from definitions based on ownership and sharing criteria. I have chosen to go with the latter, for reasons that should become evident.

D3. Sharing. "Sharing" refers to whether the facilities used by a network are dedicated to a particular user or shared by other users. The dedicated/shared dichotomy is to transmission capacity as the open/closed dichotomy is to user groups. Dedicated lines represent circuit capacity that is "closed" to all but one user, whereas switched or non-dedicated facilities are "open"

to use by all. Dedicated facilities are the building blocks of private networks. They represent capacity that is largely under the user's control.

While the use of dedicated facilities is strongly associated with private networks, it does not seem to be sufficient to qualify a network as private by itself. Many companies with high volumes of traffic order dedicated facilities to connect their phones and computers into the public network. I would not be comfortable with classifying any company that orders a T-1 from a LEC as possessing a "private network."

Clearly, ownership, access, and sharing are all critical aspects of use-privatization. But, although all three are associated with the public-private distinction in important ways, none of them can be used by itself as the standard for making a classification.

In the next section I draw on the theory of the firm to define a new principle for making the public-private distinction. This definition incorporates the relevant aspects of ownership, access, and sharing but establishes a more general criterion which resolves the ambiguities of the other methods.

2. ANOTHER DEFINITION

The crux of the public/private distinction is that in private networks, a firm takes over the network management function for itself. Although it may order facilities, service, and equipment from outside suppliers, the real responsibility for assembling and operating a network is internalized. Instead of purchasing telecommunications service as an end user, the firm itself combines intermediate inputs into a final product.

The theory of the firm outlines several reasons why firms might rely on themselves instead of a market transaction to supply a needed service. None of them appear to fit the case of networks very well, however. The aspect of the theory that comes closest to our needs is that of asset-specificity, wherein the network can be viewed as a highly specialized product customized to the needs of a specific user. The theory's explanation of why firms exert direct control over custom-made inputs focuses on vertical integration as a way of avoiding opportunistic behavior on the part of a supplier. In the case of telecommunications networks, however, protection against opportunism does not appear to be the main consideration; rather, it is the telephone company's inability to offer sufficiently specialized products. Most telephone companies have proven to be unwilling or unable to create optimal combinations of intermediate inputs suited to the specific needs of firms.

Many firms have discovered that they can reap substantial economic and strategic benefits by taking over this function for themselves.

Thus we have a new standard for distinguishing private and public networks:

D4: Internalization of network management. This standard allows for variation in ownership, access, and sharing but retains a core distinction between "public" and "private." The issue is who assumes the management role. If responsibility and control reside within the firm, then it is a private network. If responsibility and control reside outside the firm, and the service is simply consumed as a final product, then it is not a private network.

3. OPERATIONALIZATION OF THE DEFINITION.

While this should clarify the general theory underlying my approach, the definition still needs to be operationalized. Basically, the study will classify as "private":

O1. All telecommunications facilities owned by someone other than a certified common carrier. Usage is confined to the owner and is not shared or aggregated on a commercial basis.

O2. All dedicated facilities leased from carriers, with the important proviso that the leasing firm's control must extend across both ends of a communications channel. Hence, a dedicated T-1 connecting a firm to an interexchange carrier's POP to deliver traffic into the public network would not be classified as part of a private network. The same dedicated T-1 connecting two privately-owned PBXs would be classified as part of a private network. (See Diagrams 1-3)

03. Note that this definition is neutral with respect to whether the private network is "open" or "closed." As a practical matter, most closed user-group networks are based on privately owned or dedicated facilities, so the definition classifies most closed user-group networking as private. However, the definition would not include closed networks established without the use of private or dedicated facilities.

4. TRENDS AND ESTIMATES

Stigler's (1951) and Williamson's (1975) theory of the life cycle of the firm posits a general trend toward vertical disintegration as an industry grows. Does internalization of network management represent more integration or more disintegration? There is an apparent paradox here. The vertical disintegration of the public network is accompanied by a growing number of firms integrating forward into the assembly and management of networks. Management-internalization thus could be seen as a reversion to a less specialized market structure, because functions that once were hired out are now being supplied internally. At the same time, use privatization goes hand-in hand with the disintegration of the telecommunications system and the growth of pluralism and specialization. If multiple firms can make decisions regarding intermediate network inputs independently, then multiple suppliers, competition, technological variation, and specialization in the intermediate market will be supported. The privatization of management feeds upon and reinforces the breakup of the end-to-end network into its component parts, and encourages

proliferation of the suppliers of those component parts. Hence in the long run privatization leads to much greater specialization in telecommunications services and facilities even though it internalizes or integrates the network management function.

This stimulus to specialization may eventually lead to re-concentration of the network management function into specialized suppliers, and a decline in private networking (as defined here). By this I mean that once "internalization of management" is accepted as the distinction between public and private networks there is no reason to assume a priori that the (network) "private sector" will continue to grow at the expense of the "public sector" indefinitely. Internalization may be simply a temporary response to the inadequacies of the POTS-based telcos and the pricing distortions of regulated monopoly. Network management functions that are now internalized may migrate back to third-party providers as the telecommunications industry becomes more competitive, flexible, integrated, and experienced in the ways of telematics. CitiBank doesn't want to be one of the world's biggest telecommunications companies; it has to be because no existing third-party firm has the geographic scope or expertise needed to handle its special communications needs any better than it handles them itself. Most private

networks have emerged to meet new or extremely specialized communications needs.

Our survey uncovered several indications that a trend toward re-specialization of the network management function is possible. We see it, first, in the emergence of large-scale information services platforms, and second, in the adoption of Software-Defined Networks.

Our survey identified three large-scale information service platforms (ISPs). One specializes in 800 and 900 number call processing and concentrates as many as 10,000 circuits into a single center. Another specializes in remote computing services for corporations and runs 8 DS3 pipes (a total of approx. 4600 voice circuits) into the PSN. The third provides voice, fax, and data messaging and operator-assisted services. (ISP Architecture chart) ISPs combine the sharing efficiencies of the PSN with the kind of specialized features and functions that used to require a private network. Another interesting aspect is that all three of these firms are fairly new. The oldest is a little under three years old, the newest is less than a year old. These three ISPs alone accounted for almost 60% of the local access circuit capacity documented in the survey.

Seven of the thirty large users we surveyed were using a "Software Defined Network" service for interexchange

telecommunications. SDNs provide "bandwidth on demand" and the "feeling" of a dedicated network without actually requiring dedicated circuits or internal management by the firm of its bandwidth changes.

The area of most rapid growth in private networking is the LAN. With the construction of an internal LAN, businesses, schools, government agencies and hospitals establish closed user groups and unambiguously cross over the boundary line from network consumer to network owner and manager. Our survey of 30 organizations uncovered 102 different LANs linking a total of 4,685 terminals. Nearly all LANs are less than 5 years old. Only 8 of the respondents (27%) did not have LANs. LANs were the second-most frequently designated area of expected high growth (next to high-speed data communications).

In assessing private ownership of telecommunications, it is clear that noncarrier-owned transmission facilities that extend outside an organization's own buildings or property are still ~~fairly~~ rare. Only 8 organizations owned private communications access facilities (not counting PBXs). Only four organizations owned private telecommunications access facilities that extended beyond their own building or campus. Of those, two simply owned a single link between two locations (in one case a fiber cable, in the other a microwave hop). Only two large users, then, had networked

private facilities. One was a railroad which had laid fiber along its rights of way. The other employed VSAT facilities. The railroad corporation's classification as a "private network" is dubious, however, because it has begun to provide bypass services to other users on these facilities.

5. MEASUREMENT STANDARDS

It would be nice to be able to say something to the effect of "30% of all networking in the surveyed city is private, and the proportion is growing (or declining) by 3% per year." Our survey cannot provide such an estimate, however. To begin with, it is not clear what kind of measurement standard could be used to come up with such a unidimensional figure. Several possibilities suggest themselves:

M1: Investment in facilities. This standard would count or estimate the value of the hardware invested in public and private systems. This is the sort of approach taken by Crandall (1991), which called attention to the growing amount of investment that is not made by telephone companies.

M2. Traffic (minutes and volume). Another standard would be to attempt to measure the amount of traffic that flows through the public and private systems. This approach is valuable if it can be done, but faces major practical obstacles. How does one find out how much traffic is handled internally by corporate PBXs? by a university LAN? Does one aggregate or separate data, voice, and video traffic?

Information Services Platform
Architecture

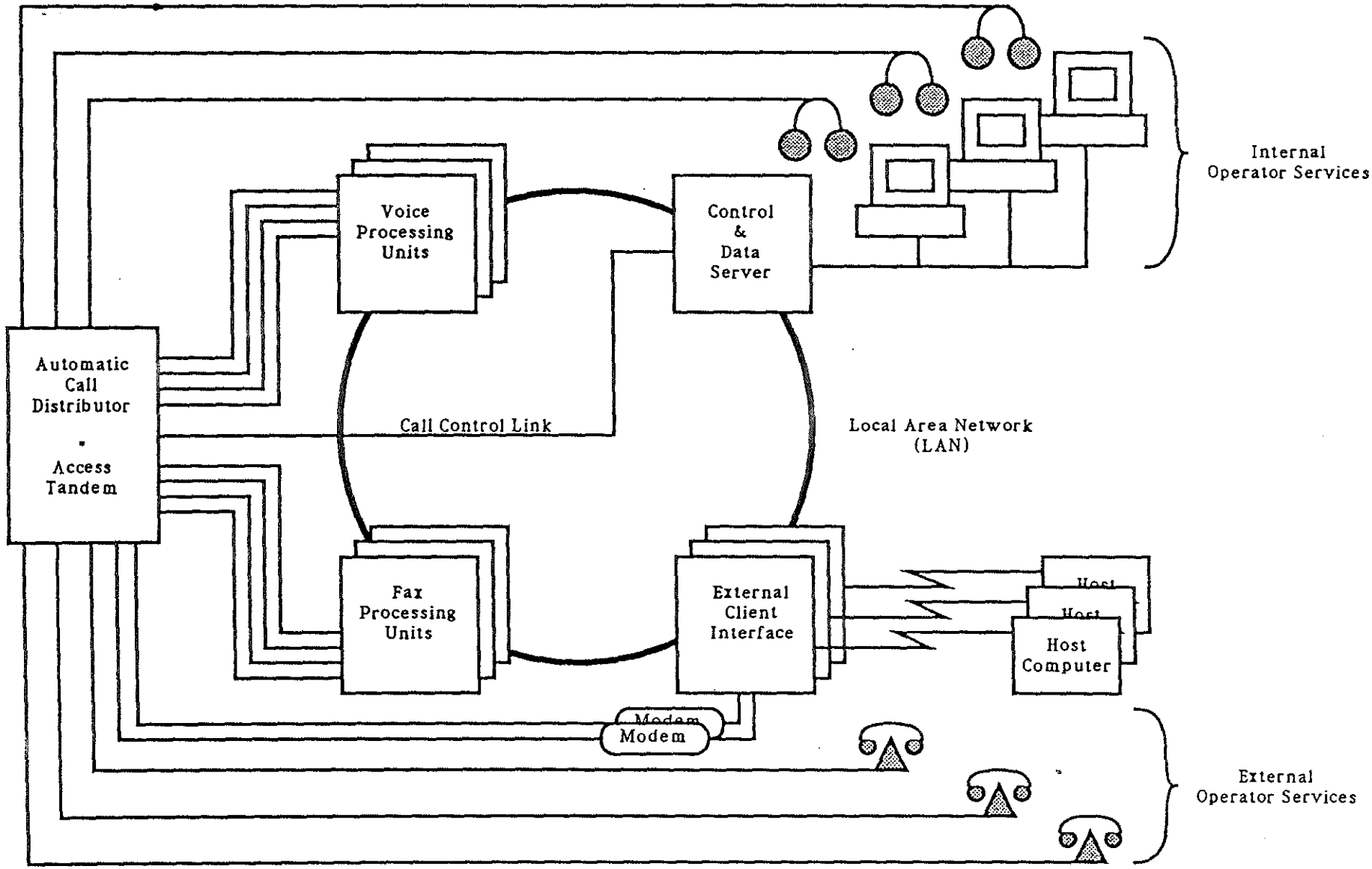

Figure 1