Network Utilization Principles and Pricing Strategies for Network Reliability

Raymond W. Lawton

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

DISCUSSION DRAFT ONLY

NETWORK UTILIZATION PRINCIPLES AND PRICING STRATEGIES FOR NETWORK RELIABILITY

by

Raymond W. Lawton, Ph.D.
Associate Director
The National Regulatory Research Institute
Adjunct Associate Professor
School of Public Policy and Management
The Ohio State University

Quality Reliability of Telecommunications
Infrastructure: The Policy Debate

Columbia Institute of Tele-Information
Columbia University Graduate School of Business
Columbia Business School

April 23, 1993

Introduction[1]

Networks are widely used, but are not fully understood. The airline, trucking, telecommunications, and computer industries are the most visible examples of the pervasive and successful use of networks in modern economies. In each of these industries the design, management, operating characteristics, and output of the networks has been extensively studied. Yet, the economic rationale of these networks, namely, why networks are used versus other modes of production, has not been adequately examined. This shortcoming is particularly significant because it is the economic advantage provided by a network that is the reason for its pervasive use and not its other characteristics.

It will be argued here that costing and pricing approaches that appear to work well for some types of networks do not seem to be well suited for other types of networks. Pricing based on short-term marginal costs, for example, produce efficient outcomes in increasing cost electric utility networks, but does not seem to be as efficient for the declining cost public switched telecommunications network. Costing and pricing schemes are generally unsuccessful when they do not accurately reflect the underlying cost structures of the network as well as the pricing environment of the network. If networks can be viewed as being either in regulated, transitional, or competitive markets, then the optimal network costing and pricing approaches to be used for each specific type of regulated, transitional, or competitive market are those that are also congruent with the underlying cost structure. A simple example of this would be the inappropriateness of using the average cost pricing long associated with regulated telecommunications markets for a fully competitive telecommunications market.

In the following section certain network utilization principles are developed based upon the costing and pricing characteristics of telecommunications networks in

---

[1] The views and opinions expressed herein are not necessarily those of The National Regulatory Research Institute, The Ohio State University or any particular state regulatory commission.

competitive markets. These network utilization principles then are able to describe costing and pricing approaches that are appropriate for certain kinds of network market environments. Since reliability is a cost factor and has measurable parameters, these network utilization principles can be applied to assist network owners or managers in designing reliability pricing schemes. Accordingly, the third section of the paper identifies reliability pricing options for competitive, transitional, and regulated telecommunications networks.[2]

Network Utilization Principles

The fundamental assumption underlying network utilization principles described below is that the construction, operation, and use of any network is based upon economically rational decisions. That is, based on the information available to a firm, a network is the least expensive means of delivering to the firm's customers a specific service. If the network is not the least expensive means of delivering a particular service, then an alternate means of delivering the service will be used or established by the firm. The assumption of economic rationality is key to understanding both the purpose served by and the utilization of a network. Unless other reasons exist that override the firm's interests in profit maximization, a network must be the low-cost service provider or it will not be used.

A network may be either a physically interconnected, ubiquitous distribution system or, an integrated system of switches or nodes and routes or channels with usage restrictions and enforceable interconnection agreements. The local distribution arrangement of electric, gas, and water utilities are generally thought to be the classic examples of the first definition, whereas the public switched telecommunications network, Peter Huber's geodesic telecommunications network, and various intermodal

---

[2] In addition to the assumption of an economically rational firm, this analysis is based upon two other important assumptions. The first is that the networks described are unregulated networks. The second assumption is of a uniform availability and usage of the same telecommunications technology and industrial organizational structure.

transportation systems are examples of the second. More specifically, a telecommunications network may be defined as a ubiquitous and economically efficient set of switched communications flows.

Ubiquity is one of two indispensable components of a network. An information or communications distribution system that does not have the facilities or instant access to facilities needed to serve all customers desiring service within an area is neither ubiquitous nor a network.[3] A telecommunications firm able only to serve a fraction of the customers within an area would not be considered ubiquitous or a network. On the other hand, a telco that provided telecommunications services to all customers within on building would be considered as having a ubiquitous network, when the building is chosen as the unit of analysis. Ubiquity, not size, is one key factor when defining a network.

An entity wishing only to link certain points or customers together is not ubiquitous and is best thought of as a point-to-point, or subnetwork. Many examples of subnetworks exist. Railroads, for example, have developed extensive private telecommunications systems that allow switched communications flows between any railroad company facility without having to use any part of the public switched network. However large the resulting railroad telecommunications system, such a system is not ubiquitous as it does not serve (or intend to serve) all potential customers within a geographical area.[4]

---

[3] A company does not necessarily have to serve every possible customer, but rather must have the facilities-based ability to link all firms desiring to be customers in a given geographic area upon request.

[4] A business located just outside the railroad company property must still be served by another telephone network. If the unit of comparison is limited to railroad facilities, then this would constitute a network since it is ubiquitous. If the unit of analysis is a region, state, or nation, the railroad instance could only be considered to be a subnetwork. A point-to-point subnetwork is also economically efficient, otherwise an economically rational actor would not have built the system and could have found it more efficient to continue to use the ubiquitous network.

In understanding network utilization the type of ownership and form of regulation are not as important as determining whether or not a network is ubiquitous. Once the major pattern of network utilization has been established, then the ownership and regulatory status of the network become significant factors affecting network utilization.

From both an economic perspective and a public policy viewpoint both ubiquitous and subnetworks are important. Even with changes in technology, market structure, perceived and actual customer demands, and regulatory policy, the future telecommunications network will include both types of networks. To the extent that the point-to-point subnetworks do not desire, need, or demand ubiquity they should continue to exist as viable nonubiquitous communications systems. No second class status is inferred by this classification or presumption that interconnection to a network is required, only that subnetworks are not ubiquitous and that this characteristic has consequences.

Ubiquity suggests another aspect of networks: namely that they are indivisible in terms of facilities and availability of communications services. Unlike a point-to-point subnetwork, a network must offer a uniform availability of a standard level of communications services for all customers in order for it to be ubiquitous. This point is developed more fully below.

The second indispensable part of a network is the economic advantage the network offers the firm using the network over other networks and over alternative modes of delivering telecommunications services. A firm will use a particular network if it is less expensive, for a given quality and type of service, than the choices available from other network, subnetwork, or nonnetwork providers. Based upon these underlying network characteristics eight network utilization principles can be drawn.

> *Network Utilization Principle Number One: A network will only be used by a firm if it is the least costly alternative for the delivery of a particular set of telecommunications services.*

The network utilization principles identified can be used for any type of network and are elaborated here using the provisioning of switched telecommunications services in a specified geographical area. The first principle directly and immediately follows from the economic advantage characteristic inherent in all successful networks. A firm using a telecommunications network explicitly does so because, for example, it has determined that a targeted fax-based sales campaign is less costly and more effective than other alternatives such as mass mailings, radio or television commercials, newspaper advertisements, or electronic bulletin boards. Using the particular telecommunications network gives the firm an economic advantage. Unless use of the network is restricted, this economic advantage is equally available commercially to all similar firms.

> *Network Utilization Principle Number Two: A firm will build, rent, or otherwise obtain its own facilities-based network when doing so is less costly than the use of existing commercially available networks.*

The first two principles are founded on the notion that an economically rational firm will choose the least costly means for the reliable delivery of its telecommunications-dependent services to its customers. Economics and engineering economics texts recognize this in their treatment of capital investment decision making by firms. As firms are interested in maximizing future revenue streams, they are indifferent as to whose facility provides them with the needed service. They care only that the facility or network chosen is the least costly alternative for a certain quantity, type, and quality of service.

A firm is, therefore, willing to consider all reasonable alternatives: using an existing network, using one or more point-to-point subnetworks, constructing its own network, or using nonnetwork providers. The bottom line here is that the alternative chosen is the least costly of all those available to the firm that will allow the firm to deliver certain services to its customers. Accordingly, if a particular network is not the lowest cost means of delivering ubiquitous service, then a second network will be constructed and all customers of that service will prefer the second network because of its lower cost. Therefore, the economically successful existence of a network is proof of principle number three.

> *Network Utilization Principle Number Three: A network is the least cost alternative for the ubiquitous delivery of certain telecommunications services.*

This principle is derived from the twin notions that only networks can provide ubiquitous service and that a firm will only use a network if it is the least costly option. A geographical area may have more than one network, and each network could have essentially identical costs.[5] From an engineering perspective, switched telecommunications networks gain their efficiency by determining the optimal configuration of switches and lines needed for the telecommunications demand pattern existing for a self-selected service area.[6] By being ubiquitous a network has a larger volume of traffic than a subnetwork and can better design economically efficient network facilities. Said another way, while the network does not know if a particular firm will use a specific service at a given time and location, the network designer does know the basic underlying aggregate demand pattern and can build facilities that can handle the demand within a known margin of error. The multi-directional star typology of public switched networks, for instance, is ideal for efficiently handling large volumes of telecommunications traffic because the trunk lines that link switches can be used for multiple purposes. As is show in figure one, path AB serves not only as the path between switch A and B but also as an alternative route for switch paths AC, AD, BC, and BD. This and other engineering optimization techniques, such as the recent development of "self-healing" fiber ring topologies help make telecommunication's

---

[5] Markets with competing networks can be stable or experience successes and failures even when using the same technology and type of industrial organization, design, or management.

[6] Lines can be wire-based, coaxial or fiber or wireless radio options such as cellular, personal communications systems, or other radio spectrum-based options.

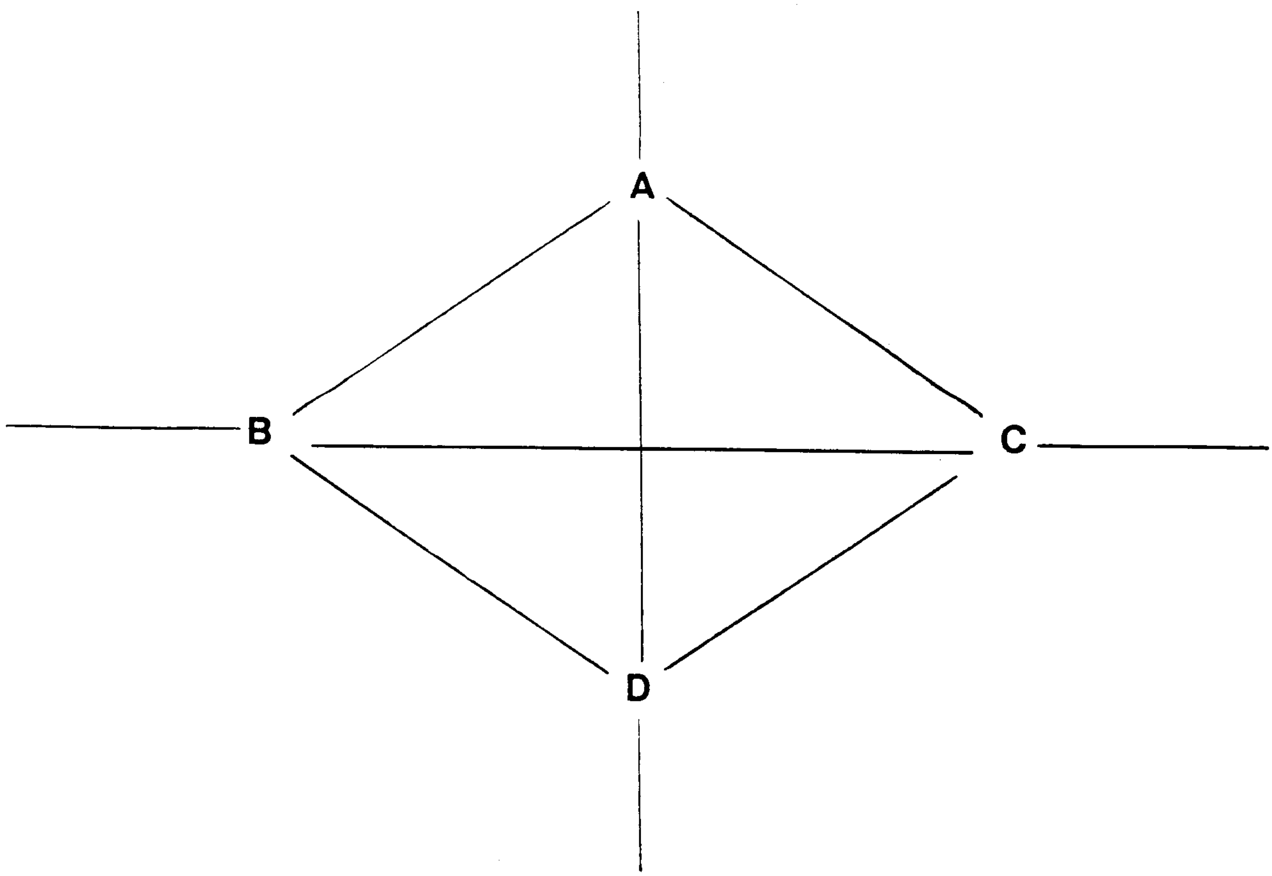

Figure 1: Illustrative network showing four switches and lines.

network efficient providers. It is important to note that networks are not always the least cost alternative, only that use of a network by a firm means (that based upon the information available) a firm has selected a particular network as the lowest cost means of delivering or receiving certain services.

> *Network Utilization Principle Number Four: A point-to-point or subnetwork is the least cost alternative for the nonubiquitous delivery of certain telecommunications services.*

An economically rational actor will not build a subnetwork unless doing so will increase its net revenues and its profits. In general, a point-to-point subnetwork is viable because it has a different underlying economic structure than that of the ubiquitous network. Three important elements describing this difference are (1) traffic concentration, (2) the cost of ubiquity, and (3) the nonubiquitous nature of the services to be sold. In consideration of the first point, many networks have a channel or set of channels that carry a disproportionate share of the traffic of the network. The current customers of a given network will consider the construction or rental of their own subnetwork if they do not need ubiquitous service and instead need only the high volume routes of the incumbent network. A firm will look for a subnetwork if the change will minimize costs, and the users do not require ubiquitous service. Retail store chains with their own internal telecommunications systems are good examples of successful point-to-point subnetworks, although they still depend on the ubiquitous networks for their external calls.

Second, while the primary advantage of a network is its ubiquity, its greatest structural weakness--namely, the presence of localized cost diseconomies--is due to the ubiquitous nature of the network. Just as significant economies are realized by the strategic inclusion of high traffic routes into a network, substantial diseconomies occur when all customers in a geographical area are linked by the network.[7] The net result of these diseconomies is to provide incentives to the customers needing only to use certain routes to consider the development of a subnetwork that minimizes their communications costs.[8]

---

[7] Diseconomies occur because the cost of low traffic channels on a per call basis is greater than the average cost of all calls on the network.

[8] The analogy to be drawn here is that the cost structures that produce the economic efficiency of the network are the centripetal forces that keep the network whole and cohesive. The diseconomies present in portions of the network that do not have uniform costs and traffic flows are the centrifugal forces that can cause the network to be reconstituted. Absent any other constraints or goals, an economically rational firm weighs the net of these centrifugal and centripetal forces and either stays on the network, or participates in some form of a subnetwork or nonnetwork mode. A network with ubiquitous service has a different economic structure than a subnetwork. That is, subnetworks tend to be more uniform, being composed, for example, of high volume routes only.

Third, a point-to-point network does not, by definition, provide by itself ubiquitous service.[9] Unless a previous user of a network never used the network for anything other than the specialized services now provided by the subnetwork, some connection will likely be maintained with a network. However, to the extent that a nonubiquitous service can be provided more economically than had been provided by the network, a point-to-point subnetwork will be preferred.

Networks and subnetworks can and do coexist and prosper when each fills a different need. All other things being equal, subnetworks can only come into being when a network is not the lowest cost alternative. Again, the economic existence of such a dual telecommunications system is by itself proof, absent any market imperfections, of the need for such a dual system.

Network Utilization Principle Number Five: All services use the network in order to obtain the network subsidy.

A network is used for only one reason: it is cheaper than any other alternative. Accordingly, use of the network by a firm effectively constitutes a subsidy that equals the difference between what an alternative network would cost the firm--that is, its annualized cost--minus the annual price charged to the firm for use of an existing commercially available telecommunications network. A freely negotiated contract between the firm and the owners of the network would result in a charged price that reflected the comparative value of this subsidy to both parties.[10]

---

[9] By various contractual means subnetworks may be linked with networks or to other subnetworks. In a system where prices follow costs the subnetwork will pay its fair share of network cost. Recall, however, that the continued existence of a subnetwork proves it is economically viable for whatever specialized purpose it is put to be it owners and customers. It is only when the subnetwork desires additional or ubiquitous services that it seeks some form of network access.

[10] The value placed on the subsidy price agreed to by the owner of the network is the end result of an optimization process that weighs and evaluates the preferences and risk-taking style of the network owner and the firm. It also involves judgements about the

As long as the subsidy stays within the preference range of the firm--which is between the network short-run marginal cost and the annualized cost of self-provisioning--it will stay on the network. In a system where all users of the system have perfect information about the cost of alternative networks, subnetworks, and nonnetworks for all other parties, the size of the subsidy received by each party will be known and agreed upon within some margin of tolerance. In regulated public-switched and telecommunications networks, government agencies or legislatures establish policies that authoritatively determine the subsidy enjoyed by each customer.

If every user of a network had an identical usage pattern and similar location, then costs would be identical for each user and, presumably, the prices charged would be identical. Because the use of a network is rarely exactly uniform by the customers of the network, costs incurred and prices charged also tend to vary by customer.[11] Armed with this knowledge, the network's actual and potential customers have a twin set of incentives. The first is to ensure that their firm pays the lowest possible price. The second is to ensure that the price charged other customers does not cause the price charged to their firm to be higher than that expected to be achieved by negotiation. Discipline is enforced in any price-setting negotiations by Network Utilization Principle Numbers One and Two.

The agreed-upon price should be between the cost to the buyer of obtaining or constructing an alternative network and the cost to the owner of the network of providing the service. The difference between the cost-based prices charged that a customer would have to assess itself to recover the cost of using or building an

---

elasticity of demand for the services provided by the network and an appraisal of the availability of substitutes.

[11] Recall it is assumed that prices charged are cost-based. Without this assumption a network may choose to offer, or be ordered to offer by a governmental agency, a price to a customer that does not necessarily recover the cost of serving that customer. Lifeline rates and rates charged to the hearing-impaired are examples of rates that do not cover the cost of providing the service. In regulated systems higher prices are charged on the nonhearing-impaired services sold to recover any loss. In unregulated systems the owner of the network may pursue a variety of strategies to cover any such self-imposed losses.

alternative network and those ultimately agreed to by the customer and the network owner constitute the network subsidy. If no subsidy is available, either because (1) the network owner posts prices that are above the prices charged by alternative networks, or (2) the legitimate cost of the network is higher than available subnetwork or nonnetwork substitutes, then the potential customer should build or rent an alternative network, or abandon its planned sale of services. The network subsidy is a natural by-product of and occurs in any decreasing-cost network. The subsidy is available, in principle, to all customers and is the product of any economies of scale and scope that the network enjoys.[12]

The ability of an actual or potential customer to obtain some or all of the subsidy is affected by a number of factors. These include, but are not limited to, the information available and the availability of substitutes. Both of these factors directly affect the bargaining power and ultimate outcome of negotiations. In a monopoly situation with no technologically equivalent substitutes available, the bargaining power of the firm is inferior to that of the network owner. In other market situations the relative bargaining power of firm and network change accordingly.

It is important to recognize that the network subsidy is not necessarily a cross-subsidy. Generally a cross-subsidy is said to exist when one service has its price explicitly set above its incremental cost so that the price for another service can be set below its incremental cost (Berg and Weisman, 1991). In the case of a declining cost ubiquitous network the highest possible price a network can charge a firm is equal to or below the firm's self-provisioning cost. Because of this fundamental feature, less incentive exists to charge prices below the network's short-run marginal cost as the firm has already decided that the network is the least cost alternative. The only issue to the firm is how much of the network subsidy it can obtain. A cross subsidy in an unregulated network is

[12] It is difficult to know exactly what the size and scope of the economies enjoyed by a network are. To do so would require the accurate and complete costing of a nearly infinite variety of alternative networks and point-to-point subnetworks. For all practical purposes it seems sufficient to assume that the continued economic viability of a network is enough proof that a network has positive but unknown economies.

only possible when significant information asymmetries exist or when the negotiating firms agree to provide a cross subsidy when markets are imperfect.[13] The key pricing issue for the network owner is how to distribute the subsidy in such a way that customers and profits are not lost.

Network Utilization Principle Number Six: A network is integrated and indivisible.

Unlike some other modes of industrial production, such as those found in retail stores, insurance, and the agriculture sector, a network has as its core reason for existence the ability to uniformly and automatically connect all of its customers with each other upon demand. While some routes, services, or switches may be more heavily used than other routes, services, and switches, a pattern of differential network usage does not affect the fundamental connectivity that all networks sell. The need to achieve ubiquity for a self-defined area combined with the need to employ engineering optimization strategies in order to achieve cost advantages over self-supply options and alternative nonnetwork-based competitors have the combined practical effect of making a network indivisible and integrated.

If an existing telecommunications network was broken into its constituent parts, the result would be either a set of point-to-point subnetworks or a set of nonoverlapping, but smaller networks. The smaller networks, however, would still be indivisible and integrated, but for a smaller geographical area. Once ubiquity is lost, a network loses its fundamental ability to connect all of its former customers. It is no longer integrated or indivisible. There is no economic-based presumption that subnetworks should become networks or that the disaggregation of existing networks into subnetworks is something to be avoided. Instead, the logic underlying the network utilization principles suggests that economic demand will be the initial basis for determining whether a firm needs a

---

[13] It does not matter what costing method the firm or network employs as long as they believe the information to be sufficiently accurate and subsequent network utilization decisions are based upon this information.

network, or a subnetwork, or a nonnetwork. It is the aggregate pattern of the demand for ubiquitous and nonubiquitous telecommunications services of all firms that determines the number, size and mix of telecommunications networks and point-to-point subnetworks.

To examine this point further, it is important to ask if an optimally designed network that had the most efficient possible arrangement of lines and switching could be reliably disaggregated into its constituent parts for costing and pricing purposes. Imagine two central offices, A and B, linked to each other through a trunk connection. Could the customers of central office A convincingly argue that their costs are only the costs for central office A and not for the trunk line connecting the two offices? In this instance a trunk connection would not be built if sufficient demand from the customers of A and B did not exist. The linking of the two central offices indicates that the "local" ubiquity of the two previously stand-alone central offices has been replaced by the ubiquity represented by the combined A-B network. The size of the trunk connector simply represents the calling volume expected within specified technology and quality of service parameters. It is the presence of the connecting trunk itself that indicates the demand to expand the ubiquity of the telecommunications network. If only local ubiquity is desired by a critical number of firms then networks consisting only of central offices A and B could be operated. To the extent, however, that the net increased demand for trunk A tandem-based services provides revenues that allow a continued A-B network to price its "local" services and allocate its network subsidy such that the prices for local are lower than stand-alone central office A or B, then the expanded network should prevail. To the extent that the increased ubiquity is not desired, or is not economically efficient, or does not produce lower prices, then the enlarged A-B network should fail.

This form of reasoning can be extended and generalized to deal with any network disaggregation proposal. If the decision to build facilities is irrational from an economic perspective, the customers will choose a least cost alternative that meets their communications needs. What tends to make the network attractive to a firm is the traffic optimization capabilities of the network. A network achieves traffic optimization because of the "law of large numbers." The large customer base of the network enables it to build less capacity because while every customer is connected to the network, not every customer has the same communication profile. Accordingly, a network can build a proportionately smaller network than might be built by a firm having a lower amount of traffic.

In order to further examine the indivisibility of a network, imagine the following instance. A firm desires that the ubiquitous network provide a service consisting only of a single unswitched line to a computer owned by another firm across the street. Could the stand-alone costs for this service be reliably identified and would these costs be a valid representation of the costs incurred by the ubiquitous network? These questions can be answered by using the network utilization principles identified previously.

A firm seeking to use the ubiquitous network's facilities, or to have the network construct facilities seeks to use the network because the alternatives available to the firm are more costly. The network is able to provide the service because it has right-of-ways, along with repair, service, and maintenance facilities throughout its service territory. It has a sales, marketing, construction, and network operating ability that, within known error and quality of service standards, is equally available to all customers within the area served by the network. The price charged to the firm seeking only an unswitched, dedicated line must include an appropriate share of what in aggregate it cost the network owners to attain the cost structure that enabled the network to be the least cost alternative. In an unregulated market the existing customers of the ubiquitous network will monitor the prices charged other customers in order to ensure that they are being charged a fair price and receiving their fair share of the network subsidy. If the network owners decide to charge the firm requesting the single line a price less than its short-run marginal cost the existing and future customers of the network could have an incentive to seek an alternative provider.[14]

[14] Different network customers will have different perceptions about the attribution and assignment of costs and prices and will bargain accordingly. If a firm or group of firms possess market power, they may be able to obtain network services at a lower price than would otherwise be the case without their having market power. It is here that the existence and allocation of the network subsidy becomes important. As long as a firm

As noted above, one of the major reasons for the cost advantage enjoyed by a network over alternative means of providing the same service functionality lies in the traffic routing capabilities of the network.[15]

Application of Network Utilization Principles to the Provisioning of Reliability for Telecommunications Networks

Average and Special Purpose Reliability

Reliability is an important and intrinsic feature of telecommunications networks. Firms and individuals choose to use networks because of the networks ability to reliably communicate with all other customers of the network. Networks do not, of course, have the ability to provide service that is always perfectly reliable. Even so-called self-healing ring-based networks do not have the ability to ensure perfect reliability as backup routes and switching can also fail. Furthermore, rings provide increased reliability, but do so at the cost of essentially duplicating facilities in a way that can provide rapid rerouting of interrupted telecommunications traffic.

If a network's customers all had the identical need for reliability, no particularly compelling reason would exist for identifying a separable cost or price for reliability. When the need for reliability is uniform, stable, and known, its pricing will mirror the network pricing parameters described earlier. As long as the price charged for use of a network service with a given degree of reliability is less than that available by self-provisioning, or from network and nonnetwork competitors, no new issues arise.

Because of the very wide array of services available from telecommunications networks, the different construction and operational costs incurred for different levels of

---

pays less than its stand-alone self-provisioning cost and more than its marginal cost, the only issue of note is how much of the network subsidy the firm will receive.

[15] If a network is unable to offer a lower cost-based price to a firm than another network or point-to-point subnetwork, it is, by definition, not the least cost provider. This distinction is important as ubiquitous networks often compete with each other and with specialize subnetworks.

reliability, and due to the different value different firms place on each service, the pricing of reliability has become an important issue. From an engineering perspective two kinds of reliability are possible. The first is average system reliability (ASR) and the second is special purpose reliability (SPR). ASR is the average ability of any part of the network to deliver uninterrupted communications upon demand to and from any part of the network. The failure or error rate is known and randomly distributed for certain types of facilities. In general, all customers of the network have the same average level of reliability.

Special purpose reliability is quite different. By way of analogy, average system reliability is to average cost pricing as special purpose reliability is to short-run marginal cost pricing. SPR occurs when, say, a fiber token ring is constructed to serve specific, generally large-volume customers. The ring typology is an efficient way to provide alternative routing or switching so that traffic that otherwise would have been interrupted can reach its intended destination. This kind of reliability is special because only a portion of the network is generally served. If the whole network had a token ring structure, then this special reliability would become average system reliability. Further, while all networks have ASR they do not necessarily have special purpose reliability.

For regulated and unregulated networks a tension exists between the network customers that desire only ASR, versus those who need special purpose reliability. In regulated public switched telecommunications networks this largely parallels the debate over plain old telephone service (POTS) and enhanced telecommunications services. In a regulated network facing little competition the network customers that require higher or special levels of reliability would be in favor of pricing policies that spread the cost of special reliability across all network customers. Here the logic is that special reliability is a network feature that could eventually be available to all network customers, and that an intertemporal shift of costs is in the public interest and the eventual interest of all network users. POTS customers view special purpose reliability as a premium service that is a private good and should be strictly paid for by the direct users of the special purpose reliability. POTS customers see no necessary externalities in efforts to achieve special purpose reliability, and desire only to pay for average system reliability. Disputes over the apportionment of the costs of average and special purpose reliability are resolved by commission rulings where networks are regulated.

Unregulated telecommunications networks in competitive environments face additional constraints. Presumably, average levels of reliability are known and are among the factors used by a firm in selecting a particular network. If the choice is between one network with one hundred percent token ring backup versus a network with no token ring backup, then the firm can sort out its price and reliability preferences and select the network having the best combination. In hybrid networks where only selected customers have token fiber ring backup the choice is more complicated but follows the same price/reliability decision-making logic. The key reliability pricing issue lies in determining why, if a nontoken ring network is equally available, a firm that is not directly serviced by the token ring chooses the particular token-ring network. As long as the special purpose reliability customers pay for their special service, no particular problem occurs. Indeed, for a hybrid network, it should normally be expected that the SPR self-provisioning costs are higher than those firms needing only ASR.

If special purpose reliability is being thought of and treated as ASR and is being phased-in and the lead time is acceptable to a nontoken ring firm, then a price/reliability balance may be achieved. It would be an unstable and unsustainable balance if an unregulated network charged a portion of the special purpose reliability costs to those who only demand average system reliability. With competitive options available, the average system reliability customers would seek a network whose prices better reflect the reliability levels actually desired and used. Only if no other networks were available, or if the cost of a self-supply option was unacceptable, of if the incumbent network had significant market power, or if the cost of a self-supply option was unacceptable, would an unregulated network be able to enforce this type of pricing. The network owner has some ASR/SPR pricing freedom when it allocates the network surplus. This freedom is constrained by the above mentioned constraints.

Modernization

Reliability is inextricably intertwined with the type of network technology. Fiber, radio-wave, and digital-switching technologies offer increased reliability over copper lines and older switching technology. To the extent that increased reliability lowers network costs, the main issue is how the network owners use the reliability-driven cost savings, and this will depend on the goals of the network owners and the competitiveness of their markets. A larger problem occurs if increasing reliability--either for average or special purpose reliability--also increases a network's costs.

In standard engineering and economics texts, modernization decisions are uniformly described and based on clear, elegant, and powerful decision rules that say that a modernization investment should be made only if it will increase net future revenue streams. Following this modernization decision rule a network will not invest in a new technology--one having a known reliability cost structure--unless doing so will profitably increase future revenue streams. Unless the modernization analysis is flawed or conditions change (say, the forecasted demand does not occur), both average system and special purpose reliability costs and resulting prices should be expected to be more favorable than those existing before the modernization investment. Otherwise those firms disadvantaged could go to an alternative network. Accordingly, if the modernization decision rule is followed only forecasting or data problems can cause problems with either average systems or special purpose reliability costs. Assuming the availability of competitive networks, if the modernization rule is not followed then the network may lose customers and may fail.

In figure two possible outcomes for networks correctly following the modernization decision rule in competitive and noncompetitive markets are displayed. Network failure (depending on the magnitude of the modernization investment) occurs most quickly in competitive markets when unsustainable and uneconomic modernization decision is made. Because a network's customers have options they can easily migrate to networks that have followed the modernization decision rule correctly. Even if the networks prices are still less than the self-supply prices for a firm, it is the availability of a competitive network that only modernizes that rationally allows the firm to migrate.

Are Competitive Networks Available? / Did the Network Use the Modernization Decision Rule for New Technology?	Yes	No
Yes	Successful outcomes for network and firm because all networks will adopt new technology and prices. Pattern of savings distribution is known.	Failure and predicted migration to networks correctly following modernization decision rules in order to obtain more favorable prices.
No	Successful outcome, but only incumbent network is available and pattern of distribution of savings from modernization unknown.	Stable situation as no migration possible. New entrants, or self-supply options will be considered unless prohibited by regulators or precluded by incumbent network's market power.

Figure 2: Possible outcomes in competitive and noncompetitive markets for alternative applications of modernization decision rules.

A stable situation occurs when no competitive options are available. Unless the self-supply is a viable option or new entrants are eventually allowed or induced in, the monopoly or otherwise noncompetitive network does not experience customer loss. It may, however, experience a decrease in demand. It is the elasticity of the demand that will determine the impact of inefficient modernization prices.

The most successful outcome occurs when competitive markets exist and the modernization decision rule is followed. Here both the network and the network's customers are better off. Because there are legitimate options readily available, the network owners do a more efficient job of distributing the increase in the network surplus due to the modernization investment. A less successful outcome also occurs when markets are not competitive but the modernization decision rule is followed. It is the unknown distribution of the modernization savings that makes this example problematic and not as desirable.

Conclusion

In practical terms, as long as network is less costly than self-supply options and if no realistic competition exists, network owners have considerable freedom in choosing their pricing strategies for any service that they sell, including reliability. As depicted below, pricing freedom decreases when options are available.

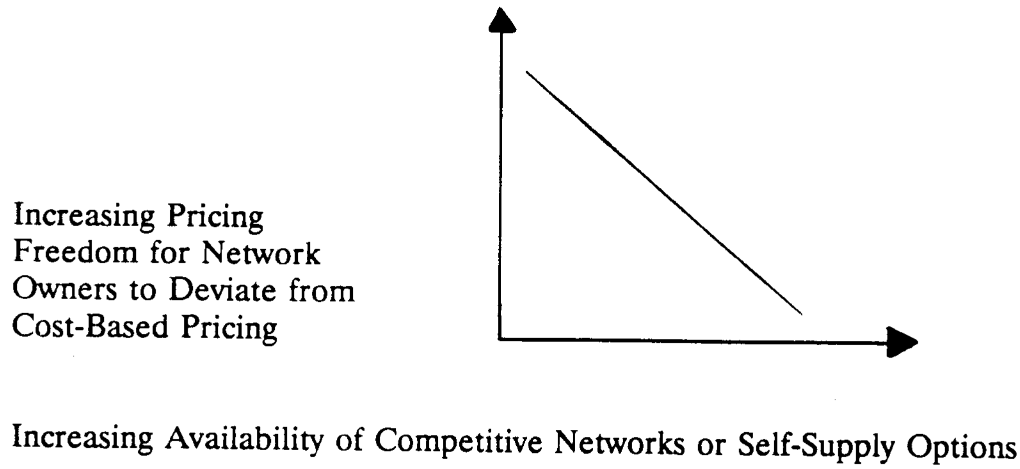

Figure 3: Network pricing freedom and the availability of competitive networks.

Deciding who pays for reliability is driven first by the network utilization principles, where a firm determines whether self-supply or the use of an externally-provided telecommunication network is in its best interest economically. The second factor is the availability of competitive networks. Reliability likely will be priced (to the degree that it is identified and priced separated) in ways that reflect the market for telecommunications networks. The traditional postal telephone and telegraph (PTT) agencies, for example, faced no significant competition and priced their services (including reliability) largely by administrative fiat. It is now widely thought that long distance toll networks are competitive enough that competitive pricing strategies (including those needed for reliability) are followed and prevalent.

Two less powerful factors affecting the pricing of reliability are the adherence to modernization decisionmaking rules and the different levels of reliability that may be present in a network. Ideally the cost of reliability should increase only when modernization rules are not followed. The more difficult issue is paying for average system reliability versus special purpose reliability. As long as average reliability is cheaper than self-supply and equivalent to other networks, no special pricing problem exists. When no competitive networks are available and self supply is not a viable option, the network owners have considerable pricing freedom when providing special purpose reliability-based services. Only when alternatives exist do network owners have to price reliability competitively.

From a public interest perspective, policies that encourage self-supply options, or the emergence of competitive networks are important for pricing reliability efficiency. The recent opening up of the Class 5 Office bottleneck to facilities-based competition is extremely important in this regard. Self-supply options should be increased in open network architecture-type unbundling approaches. As competition strengthens, these and other similar approaches should make existing and emerging networks more efficient and result in a better pairing of reliability costs and prices.