Chapter 6

Network Utilization Principles and Pricing Strategies for Network Reliability

Raymond W. Lawton
The National Regulatory Research Institute
The Ohio State University

1. INTRODUCTION

Regulatory and public policymakers have become increasingly concerned about the current and future reliability of the public-switched telecommunications network. Part of this concern is driven by the worry that the cost-efficiency incentives contained in price caps and other regulation reforms may cause networks to reduce their costs in ways that also reduce the reliability of the network. A second source of concern is more immediate and is sparked by the number of recent and widely reported outages affecting major U.S. telecommunications carriers. The third is the interest in ensuring that a modern communications infrastructure exists that can reliability serve the new and emerging needs of an information-age economy. However, reliability issues have not enjoyed the same level of attention that pricing, costing, and market structure have received from regulators, policymakers, and academic researchers. Accordingly, a widespread consensus on how to define, measure, price, or resolve network reliability is lacking.

Complicating the reliability issue even further is the notion that although networks are widely used, they are not fully understood. The airline, trucking, telecommunications, and computer industries are 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.[1] 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 as well as its reliability and other operating characteristics that account for the successful use of networks in modern economies.

It is argued here that costing and pricing approaches used by state and federal utility regulatory commissions that appear to work well for some types of networks do not seem to be as well suited for telecommunications networks. Commission-approved pricing based on incremental costs, for example, has produced visibly efficient outcomes in the electric utility networks because of their increasing-cost curve, but it does not seem to be as efficient for the declining costs of public-switched telecommunications networks. Costing and pricing schemes are generally unsuccessful when they do not accurately reflect the underlying cost structures as well as the pricing environment of the networks. If networks can be viewed as being in either regulated, transitional, or competitive markets, then the optimal commission-authorized networks' costing and pricing approach to be used for each specific type of market is the one that is also congruent with the underlying cost structure. A simple example of this would be the inappropriateness of a regulatory commission using the average cost pricing long associated with regulated telecommunications markets for a competitive telecommunications market.

In the following section certain network utilization principles are developed based on the costing and pricing characteristics of telecommunications networks in competitive markets. These network utilization principles are used to describe costing and pricing approaches that are appropriate for certain kinds of network market environments. Use of these principles allows a regulatory commission, a utility, and firms in a telecommunications market to design efficient pricing schemes. Because reliability is a key 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 this chapter uses these principles to identify reliability pricing options that regulatory commissions can use for competitive, transitional, and regulated telecommunications networks.[2]

2. NETWORK UTILIZATION PRINCIPLES

The fundamental assumption underlying network utilization principles described here is that the construction, operation, and use of any network is based on economically rational decisions. That is, based on the information available to a firm at any point in time, a network is the least expensive means of delivering to the firm's customers a specific telecommunications service or set of services. If a 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 vision of a geodesic telecommunications network,[3] and various intermodal 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 the facilities, needed to serve all customers desiring service within an area is neither ubiquitous nor a network.[4] 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 telecommunications company that provides telecommunications services to all customers within a building would be considered as having a ubiquitous network for a particular service or set of services when the building is chosen as the unit of analysis. Ubiquity, not size, is one key factor when defining a network.

In a regulatory setting ubiquity is important because a utility is awarded a franchise and must provide service on demand and for a commission-approved price to any consumer within the geographical boundaries of its service territory. For a regulated utility ubiquity is both an obligation and a compelling economic fact of life. If significant economies of scale or scope exist, then the utility may be both the low-cost provider and the only ubiquitous provider. This combination gives the incumbent-regulated telecommunications utility a significant commission-supported advantage over both ubiquitous and nonubiquitous challengers. As telecommunications markets evolve toward being competitive, regulators and policymakers need to better understand the often overlooked concept of ubiquity in order to design optimal, transitional, pricing strategies for currently regulated telecommunications markets.

An entity wishing only to link certain points or customers together is not ubiquitous and is best thought of as a point-to-point network, 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 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.[5] A local area network (LAN) or metropolitan area network (MAN) or token ring provider that chooses only to serve a particular market segment within a geographical area would also be a subnetwork.

In understanding network utilization the type of ownership and form of regulation are not as important as knowing that a network provides ubiquitous service. 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 networks 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 pricing and competitive 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, whether or not it has franchise obligations, must offer the uniform availability of a standard and reliable level of communications services for all customers in order for it to be ubiquitous. This point is developed more fully later.

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 level of reliability, quality, and type of service than the choices available from other network, subnetwork, or nonnetwork providers. Based on these underlying network characteristics, six network utilization principles can be stated. The network utilization principles identified can be used for any type of network or network service such as reliability and are elaborated here using the provisioning of switched telecommunications services in a specified geographical area.

Network Utilization Principle 1: A network will only be used by a firm if it is the least costly alternative for the delivery of a particular service or set of telecommunications services.

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 network-provisioned, 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 selected 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 2: 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, level of reliability, 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 reliably 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 of, for example, reliability. Therefore, the economically successful existence of a network is proof of principle 3.[6]

Network Utilization Principle 3: 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.[7] 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.[8] By being ubiquitous, a network has a larger volume of traffic than a subnetwork and can better design economically efficient network facilities.[9] Said another way, whereas the network manager does not know if a particular firm will use a specific service at a given time and location, the network manager does know the basic underlying aggregate demand pattern and can build facilities that can handle the demand within a known margin of error. The multidirectional star typology of public-switched, regulated networks, for instance, is ideal for efficiently and reliably handling large volumes of telecommunications traffic because the trunk lines that link switches can be used for multiple purposes. This and other engineering optimization techniques, such as the recent development of "self-healing" fiber-ring typologies, help make telecommunication networks 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 (based on the information available) means a firm has selected a particular network as the lowest cost means of delivering or receiving certain services.[10]

Network Utilization Principle 4: A point-to-point network or subnetwork is the least cost alternative for the nonubiquitous delivery of certain telecommunications services.

An economically rational actor will not build or use 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 traffic concentration, the cost of ubiquity, and 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. Unless given price discounts, customers of a 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, maintain the same level of reliability, and the users will 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 a ubiquitous public-switched network for their external calls.

Second, although 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.[11] The net result of these diseconomies is to provide incentives to the firm needing only to use certain routes to consider the development or use of a subnetwork that minimizes its communications costs.[12]

Third, a point-to-point subnetwork does not, by definition, provide by itself ubiquitous service.[13] Unless a firm never used the ubiquitous 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 for a firm. 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 5: All services use the network in order to obtain the network surplus.

A network is used for only one reason: It is cheaper than any other alternative. Accordingly, the price charged to a firm by a network is influenced by the size and availability of the network surplus. The surplus is equal to the difference between what the self-provisioning of an alternative network would cost the firm—that is, its annualized incremental cost—minus the annualized incremental cost incurred by an existing commercially available network. The firm does not know the size of the network surplus as only the network owners or managers know its demand-adjusted incremental cost. All the firm knows for certain is its stand-alone cost and the announced price of the network.

The network, however, does know where its announced price to the firm is in relation to its incremental cost.[14] A cross-subsidy occurs when the price charged is below the network's incremental cost. Unless otherwise constrained by competition, market structure, or regulation, the network has the pricing flexibility to give a firm a price anywhere between its incremental cost floor and the stand-alone cost of the firm. The firm and the network have imperfect knowledge about each other's costs. A freely negotiated contract between the firm and the owners of the network indicates the comparative value of this surplus to both parties.[15] The surplus is one important source of the network's pricing flexibility.

As long as the network price stays within the preference range of the firm—which is between its understanding of the network's incremental cost and the annualized cost of self-provisioning—it will stay on the network. In a system in which 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 network surplus received by each party will be known and agreed on within some margin of tolerance. In regulated public-switched telecommunications networks, government agencies or legislatures establish policies that authoritatively determine the amount of the network surplus applied to the prices charged to each customer. As regulated telecommunications become more competitive, these policies as well as the size and recipient of the surplus will change.

If every firm using a network had an identical usage pattern, reliability needs, and a similar location, then costs would be identical for each firm 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.[16] Armed with this knowledge, the network's actual and potential customers have a twin set of incentives. The first is to ensure that its firm pays the lowest possible price. The second is to ensure that the prices charged other customers do not cause the price charged to the firm to be higher than that expected to be achieved by negotiation. Discipline is enforced in any price-setting negotiations by Network Utilization Principles 1 and 2.

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 firm would have to assess itself to recover the cost of using or building an alternative network and those ultimately agreed to by the firm and the network owner are made possible by the network surplus. If no surplus is available, and the network owner posts prices that are above the prices charged by alternative networks, or 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 network-based services. The network surplus is a natural by-product of and occurs in any decreasing-cost network. The surplus is available, in principle, to all customers and is the product of any economies of scale and scope that the network enjoys.[17] In regulated networks, government agencies determine the availability of the surplus.

The ability of a firm to obtain some or all of the network surplus 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 more competitive market situations, the relative bargaining power of firm and network change accordingly.

It is important to recognize that the network surplus 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.[18] In the case of a declining-cost, ubiquitous telecommunications 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 incremental 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 surplus it can obtain. A cross-subsidy in an unregulated network is only possible when significant information asymmetries exist or when the negotiating firms agree to provide a cross-subsidy when markets are imperfect.[19] The key pricing issue for the network owner is how to distribute the surplus in such a way that optimizes profits and customer retention.

Network Utilization Principle 6: A network is integrated and indivisible.

Unlike some other modes of industrial production, such as those found in retail stores and in the insurance and agriculture sectors, a network has as its core reason for existence the ability to uniformly, reliably, and automatically connect all of its customers with each other upon demand. Although 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 and ubiquity 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 non-network-based competitors has 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 network, or a subnetwork, or a nonnetwork. It is the aggregate pattern of the demand for ubiquitous and nonubiquitous telecommunications services 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, reliability, 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–B, tandem-based services provides revenues that allow a continued A–B network to price its "local" services and allocate its network surplus 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 could 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. This includes decreasing their demand. 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." Being able to take advantage of this "law" is one of the main reasons franchised utilities with an obligation to serve have been so successful. The large customer base of the network enables it to build proportionately less capacity because whereas every customer is connected to the network, not every customer has the same communications profile. Accordingly, a successful 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. In particular, the network of a regulated telecommunications utility is able to provide the service at a lower price because it has valuable government-approved 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 reliability 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 includes 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 surplus. If the network owners decide to charge the firm requesting the single line a price less than its incremental cost, the existing and future customers of the network could have an incentive to bargain for the same price and to seek an alternative provider if the bargaining outcome is unsatisfactory.[20]

As noted earlier, 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.[21]

3. APPLICATION OF NETWORK UTILIZATION PRINCIPLES TO THE PROVISIONING OF RELIABILITY FOR TELECOMMUNICATIONS NETWORKS

3.1 The Significance of Reliability in Telecommunications Networks

Reliability is an intrinsic feature of network and point-to-point subnetworks. State and federal regulators and international standard-setting bodies have traditionally been concerned with ensuring the reliability of the public-switched telecommunications network.[22] Regulators have established quality-of-service standards that specify the reliability minimums for a network. Typical reliability standards focus on items such as dial-tone delay, call completions, directory assistance, and interofficial transmission.

Regulators and network users have worried that the cost-control incentive of price caps and other regulatory pricing reforms may cause a network operator to invest less in reliability-increasing investments (such as might be needed for new and expensive generations of error-monitoring software) and to spend less on maintenance. Network owners have generally responded to these expressed concerns by stating that the reliability of the network is above existing standards and that this level of reliability should increase in the future. More importantly, network owners note that reliability is so inextricably related to their ability to sell ubiquitous service that it would be economically irrational for them to allow network reliability to degrade.

In order to meet the twin challenges of the information age and the emergence of competition, local exchange companies have made massive and sustained modernization investments. Modernization is generally thought to increase the reliability of the public-switched network. One unintended but inescapable consequence of the modernization is the increase in the amount and concentration of traffic on certain links or digital switches. Concentration of traffic may be the single most important way to increase the economic and engineering efficiency of a network.

Unfortunately this higher level of traffic concentration can increase the magnitude caused by a disruption, even with an increase in reliability levels. Fiber lines are economically attractive because many more calls can be handled on one fiber strand than can be done for the same amount of copper wire. A single fault or disruption, accordingly, can interrupt many more calls than may have been possible previously. The service outages in 1991–1992 for U.S. carriers—with AT&T's massive New York outage receiving the most attention—illustrates the magnitude and impact of service disruptions possible from a relatively small number of faults.

Following these outages, the Federal Communications Commission (FCC), responding to Congressional concerns, established a Network Reliability Council (NRC) in December 1991 to provide the FCC with expert advice and recommendations. One of the recommendations from the NRC was to establish a new threshold for reporting outages when 30,000 lines are affected and to refer reliability issues to appropriate industry forums for further analysis and recommendations.[23] The FCC has also ordered in its price caps proceedings certain modifications to its service quality and infrastructure reporting.[24] Concern about network reliability has also resulted in governmental action in the United Kingdom and Japan. For example, Gupta reported that in Japan tax and depreciation incentives have been provided by legislation to increase investments that improve network reliability.[25]

Insufficient attention, however, has been paid to improving the understanding of the cost structure underlying the provisioning of various levels of network reliability. Nor has sufficient attention been given to the impact of the modernization decision rules on network reliability. Both of these issues are addressed later through the application of the network utilization principles identified previously. Use of these principles allows regulators, utilities, and firms to make more efficient reliability pricing decisions. Current pricing practices may send the wrong price signal to regulators, utilities, and firms and may result in the inefficient provisioning and pricing of reliability. As telecommunications networks become increasingly complex, competitive, and technologically advanced, the successful resolution of reliability issues becomes even more important.

3.2 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, but they do so at the added cost of essentially duplicating facilities in a way that provides rapid rerouting of interrupted telecommunications traffic.[26]

If a network's customers all had the identical need for reliability, no particularly compelling reason would exist for identifying a separate 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 the 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 reliability, and because of the different value firms place on each service, the pricing of reliability has become an important issue. From an engineering perspective, two kinds of reliability are possible: average system reliability (ASR), and 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 service reliability.

Special-purpose reliability is quite different. By way of analogy, ASR is to average cost pricing as SPR is to incremental 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 ASR. Furthermore, although all networks have ASR, they do not necessarily have SPR.[27]

For regulated and unregulated networks, a tension exists between the network customers that desire only ASR versus those who need SPR. 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 and socialization of costs is in the public's interest and in the eventual interest of all network users. POTS customers view SPR as a premium service that is a private good and should be strictly paid for by the direct users of SPR. Furthermore, they see this type of averaging as running counter to the cost-causation and unbundling principles that are necessary to make a network efficient. POTS customers see no necessary externalities in efforts to achieve SPR and desire only to pay for ASR. 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 100% 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 in which 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 non-token-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 SPR 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 SPR is being thought of and treated as ASR and is being phased in and the lead time is acceptable to a non-token-ring using firm, then a price/reliability balance may be achieved. It would be an unstable and unsustainable condition if an unregulated network charged a portion of the SPR costs to those who only demand ASR. With competitive options available, the ASR 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, or 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 ASR/SPR pricing freedom when it allocates the network surplus as demonstrated through the previous examination of network utilization principles.

3.5 Modernization Decision Rule

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, which 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 if and 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 SPR costs and resulting prices should be expected to be more favorable than those existing before the modernization investment. Otherwise, as indicated by the network utilization principles, 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 SPR costs. Assuming the availability of competitive networks, if the modernization rule is not followed, then the network may lose customers and may fail.

In Fig. 6.1, 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 decisions are made. Because a network's customers have options, they can easily migrate to networks that have followed the modernization decision rule correctly.

A stable situation occurs when no competitive options are available. Unless 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, application of the network utilization principles indicates that the network owners

	Did the Network Use the Modernization Decision Rule for New Technology?	
Are Competitive Networks Available?	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.

FIG. 6.1. Possible outcomes in competitive and noncompetitive markets for alternative applications of modernization decision rules.

in this instance do a more efficient job of distributing the increase in the network surplus due to the modernization investment. The competitive nature of the market provides the necessary incentives for this successful outcome. 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.

4. CONCLUSION

In practical terms, as long as a network is less costly than reasonably available, self-supply options, and if no real competition exists, network owners have considerable freedom in choosing their pricing strategies for any service that they sell, including reliability.

Deciding who pays for reliability is determined first by the network utilization principles, in which 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 in ways that reflect the market for telecommunications networks. The traditional postal telephone and telegraph 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 so that competitive pricing strategies (including those needed for reliability) are followed and prevalent.

Two other factors affecting the pricing of reliability are the adherence to modernization decision making 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 services requiring special-purpose reliability. Only when alternative networks exist do network owners have to price reliability competitively. The basic network utilization principles can be restated in the following six reliability pricing rules:

1. Determine whether a network or subnetwork is needed.
2. Determine if self-supply or the use of an externally provided telecommunications network is the least costly alternative available to the firm.
3. Determine the availability of competitive network.
4. Determine whether the networks available to the firm follow the modernization decision rule.

5. Determine whether the firm requires average reliability or special-purpose reliability from the network for the network service it wishes to purchase.
6. Choose the pricing option that offers the lowest cost reliability for the firm among the self-provisioning, network, subnetwork, and nonnetwork options.

From a public interest perspective, policies that encourage self-supply options or the emergence of competitive networks are important for pricing reliability efficiently. 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.

The national information highway will use a number of networks and subnetworks, each having different cost and reliability parameters. Common standards can help with establishing minimum reliability floors, interconnection rules, and interoperability protocols. The network utilization principles and reliability decision rules developed here can help regulators, utility managers, and telecommunications providers develop prices that track the cost of providing the increased reliability likely to be required by the users of a national information highway.

ACKNOWLEDGMENTS

This chapter was prepared under an Ameritech Fellow award made by The Ohio State University Graduate School. The views and opinions expressed herein are solely those of the author and are not necessarily those of The National Regulatory Research Institute, Ameritech, or The Ohio State University.

ENDNOTES

1. Jean-Michel Guldmann, "Modeling Residential and Business Telecommunications Flow: A Regional Point-to-Point Approach," *Geographical Analysis* 24 no. 2 (April 1992): 121–141.
2. In addition to the assumption of an economically rational firm, this analysis is based on two other important assumptions. The first is that the networks described are unregulated networks and that competitive networks are available, as are substitute delivery mechanisms. The second assumption is that there is a uniform availability and usage of the same telecommunications technology and industrial organizational structure.
3. Peter W. Huber, *The Geodesic Network* (Washington, DC: Antitrust Division, U.S. Department of Justice, 1987).

4. 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.
5. 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 because 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 would have found it more efficient to continue to use the ubiquitous network.
6. The existence of a viable network does not mean that the network is either a natural or franchised monopoly, only that it provides ubiquitous service. A test of subadditivity to determine whether single provisioning is less costly than that by two or more providers is unnecessary here because of the sequence of decisions made by the firm seeking to buy telecommunications services. By the time network utilization principle 3 is relevant, the firm has already made decisions about its need to buy ubiquitous services and its "build versus buy" decision such that subadditivity information is irrelevant. Furthermore, subadditivity is generally not as useful in markets with competitive networks.
7. 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.
8. Lines can be wire-based, coaxial, fiber, or wireless radio options such as cellular, personal communications systems, or other radio spectrum-based options.
9. R. F. Rey, ed., *Engineering and Operations in the Bell System* (Murray Hill, NJ: Bell Laboratories, 1983).
10. Harris identified the scale, scope, network, and learning economies that occur in the public-switched network that give it significant cost and engineering advantages over alternative communications modes. These features and advantages exist in equilibrium in competitive, transitional, and monopoly markets. These advantages apply whether a firm is selecting one telecommunications service or the entire range of telecommunications service desired by the firm. Robert G. Harris, "Telecommunications Services as a Strategic Industry," in Michael A. Crew, ed., *Competition and the Regulation of Utilities* (New York: Kluver Academic Publishers, 1991).
11. Diseconomies occur because the cost of low-traffic channels on per-call basis is greater than the average per-call cost of all calls on the network.
12. 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. Subnetworks tend to be more uniform, being composed, for example, of high-volume routes only.
13. By various contractual means subnetworks may be linked with networks or to other subnetworks. In a system in which 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 by its owners and customers. It is only when the subnetwork desires additional or ubiquitous services that it seeks some form of network access.
14. Gerald R. Faulhaber, "Cross-Subsidization: Pricing in Public Enterprises," *American Economic Review* 65 no. 5 (1975): 966–977.
15. Network owners can engage in a number of pricing strategies to deal with contestable and other types of markets. See William J. Baumol, John C. Panzar, and Robert D. Willig, *Contestable Markets and the Theory of Industry Structure* (New York: Harcourt, Brace, Jovanovich, Inc.,

1982). The value placed on the surplus 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 judgments about the elasticity of demand for the services provided by the network and an appraisal of the availability of substitutes.

16. Recall that 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 service to that customer. Lifeline rates and rates charged to the hearing-impaired are examples of rates that reportedly do not cover the cost of providing the service. In regulated networks higher prices are charged, for example, on the non-hearing-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.
17. 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 an unregulated network is enough proof that a network has positive but unknown economies.
18. Faulhaber, "Cross-Subsidization"; Sanford V. Berg and Dennis L. Weisman, "A Guide to Cross-Subsidization and Price Predation: Ten Myths," Paper (November 7, 1991).
19. It does not matter what costing method the firm or network employs as long as it believes the information to be sufficiently accurate and its subsequent network utilization decisions are based on this information.
20. 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 possesses market power, it may be able to obtain network services at a lower price than would otherwise be the case without it having market power. It is here that the existence and allocation of the network surplus becomes important. As long as a firm pays less than its stand-alone, self-provisioning cost and more than its perceived incremental cost, the only issue of note is how much of the network surplus will be applied against the prices charged to the firm.
21. 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 specialized subnetworks.
22. Bruce Armstrong, *US WEST Service Quality*, Handout at the Regional Oversight Committee, Minneapolis, MN, September 1993.
23. For example, during the first quarter of 1993, some 431 outages affecting more than 30,000 lines were reported by the NRC. This compares to outages for the first quarter of 1992 and an average of 44.9 outages per quarter from January 1992 to June 1993. United States Telephone Association, *President's Report* (Washington, DC: United States Telephone Association, August 27, 1993).
24. Federal Communications Commission, *Modifications to Service Quality/Infrastructure Reporting* (Washington, DC: Federal Communications Commission, July 7, 1992).
25. Suren Gupta, *Japan Telescene*, Newsletter (Tokyo, Japan: InfoCom Research, Inc., 1993).
26. Token rings differ from the traditional star and bus network typologies normally used by public-switched telecommunications networks and cable television distribution networks. Token rings connect users via distribution panels and a line configuration that establishes a ring such that any fault or path interruption is automatically detected and traffic is instantly rerouted through a bypass switch. Operator intervention is not required because the decision to bypass and follow an alternative path is made by specific failure-detection circuits. See John D. Spragins, "Telecommunication Network Reliability Models Based on Network Structures," unpublished paper (Clemson, SC: Clemson University, 1992), p. 23; and Yael J. Assows and Vikram R. Saksena, "Economic Analysis Architectures in Two Tier Data Networks," *IEEE Network* (May 1989): 13–21.

27. The special-purpose reliability provided by a self-healing token ring is designed to produce 99.999% reliability. Richard Tomlinson, "Impact of Competition on Network Quality," Presentation at the Quality Reliability of Telecommunications Infrastructure Conference of the Columbia Institute of Tele-Information, Columbia University, Graduate School of Business, April 23, 1993.

REFERENCES

Armstrong, Bruce, *US WEST Service Quality*. Handout at the Regional Oversight Committee Meeting, Minneapolis, MN, September 1993.

Assows, Yael J. and Vikram R. Saksena. "Economic Analysis of Robust Access Network Architectures In Two Tier Data Networks." *IEEE Network* (May 1989): 13–21.

Baumol, William J., John C. Panzar, and Robert D. Willig. *Contestable Markets and the Theory of Industry Structure*. New York: Harcourt, Brace, Jovanovich, Inc., 1982.

Berg, Sanford V. and Dennis L. Weisman. "A Guide to Cross-Subsidization and Price Predation: Ten Myths." Paper. November 7, 1991.

Faulhaber, Gerald R. "Cross-Subsidization: Pricing in Public Enterprises." *American Economic Review* 65 no. 5 (1975): 966–977.

Federal Communications Commission. *Modifications to Service Quality/Infrastructure Reporting*. Washington, DC: Federal Communications Commission, July 7, 1992.

Guldmann, Jean-Michel. "Modeling Residential and Business Telecommunications Flows: A Regional Point-to-Point Approach." *Geographical Analysis* 24 no. 2 (April 1992): 121–141.

Gupta, Suren. *Japan Telescene*. Newsletter. Tokyo, Japan: InfoCom Research, Inc., 1993.

Harris, Robert G. "Telecommunications Services as a Strategic Industry." *Competition and the Regulation of Utilities*. Michael A. Crew, Editor. New York: Kluver Academic Publishers, 1991.

Huber, Peter W. *The Geodesic Network*. Washington, DC: Antitrust Division, U.S. Department of Justice, 1987.

Rey, R. F. *Engineering and Operations in the Bell System*. Murray Hill, NJ: Bell Laboratories, 1983.

Spragins, John D. "Telecommunications Network Reliability Models Based on Network Structures." Unpublished paper. Clemson, SC: Clemson University, 1992.

Tomlinson, Richard. "Impact of Competition on Network Quality." Presentation at the Quality Reliability of Telecommunications Infrastructure Conference of the Columbia Institute of Tele-Information. New York, NY: Columbia University Graduate School of Business, April 23, 1993.

United States Telephone Association. *President's Report*. Washington, DC: United States Telephone Association, August 27, 1993.