670

Interconnection and Interconnection Pricing in Local Telephone Markets

by Bridger M. Mitchell

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

PRELIMINARY DRAFT — DO NOT QUOTE OR CITE

Interconnection and Interconnection Pricing in Local Telephone Markets

Columbia Institute for Tele-Information
October 5, 1993

Bridger M. Mitchell

1.0 TYPES OF INTERCONNECTION

1.1 Interconnection of telecommunications networks comes in many varieties:

- public switched telephone networks, each of which provide end-to-end service within a geographic area. International interconnection is of this type, as well as interconnections between geographically adjacent local exchange carriers.

- mobile and portable telephone operators. These networks represent specialized types of access for end-users. For cellular operators, with exclusive licenses or limited competition in most countries, service has been in short supply (derived from limited radio spectrum). Intense user demand has supported high retail prices. In this setting the incumbent network and the mobile network have generally negotiated interconnection arrangements to mutual satisfaction. As supply is expanding through access to higher frequency bands, radio-based access increasingly threatens to substitute for some fixed-line access. The difficulties that the first digital mobile operator (Mannesmann) had in obtaining interconnection from the German incumbent network could foretell increased contention for the soon-to-emerge personal communications (PCS) operators.

- interexchange (long-distance) carriers with local exchange carriers. Largely in the long-haul transport business, US interexchange carriers require interconnection to reach most final customers. They have direct access to some large-volume customers, and have extensive retail service relationships directly with end-users.

- competitive access providers, who connect end users with interexchange carriers or other end users.

- cable television networks that provide switched narrowband service.

- data and signaling networks of end users or public network operators.

- enhanced, information, or value-added services that use access, transport, and switching facilities of a network operator.

1.2 An economic view of network interconnection

Interconnection of network suppliers is an economic relationship that has important similarities to other vertical supply relationships. It always involves a service rendered by one network operator to another network operator. That service functions as an input to the final product of the second operator. Thus, network interconnection corresponds to many other market transactions between suppliers where one firm buys a product that will be raw material or an intermediate input for the production of its own product.

Some special characteristics:

- Interconnection is an essential input to production of a call between two subscribers to different networks, and often a bottleneck input that the demander cannot produce himself.

- Interconnection requires a degree of continuing cooperation between the two operators that is unusual in ordinary purchaser-supplier relationships.

- There may be mutual (bilateral) gains from interconnection. The opening up of a second market by interconnection resembles the elimination of tariff barriers in international trade.

- The network suppliers are often unequal, in the size and value of their markets, and in the extent to which they integrate network components and services.

1.2.1 Disjoint networks

A highly-stylized view of network interconnection issues can be gained by assuming that all network use occurs withing a single period at a known uniform rate of demand. Consider a telephone call (use of network) from subscriber A in one locality N1 to subscriber B in a different area N2 (Figure 1). Connection between N1 and N2 requires conveyance of the call between the two areas. A single network that supplies both areas can be viewed as consisting of two sub-networks, N1 and N2, that are interconnected.

A call from A to B increases costs in both N1 and N2. N1 requires local access for subscriber A, call set-up, switching the call to the route to reach B, and transport of the call for its duration. In N2 the call must be switched to connect to subscriber B and connected by local access. A call from subscriber B to subscriber A incurs similar costs, except that the call setup activities are performed in network N2.

The costs of the call may be recovered by several possible retail payment schemes:

(1) caller pays. This is the widely established mechanism for directly-dialed calls.

(2) called party pays. Toll-free 800 (green) number calls and collect calls.

(3) both parties pay. Cellular subscribers pay for "air time" under most tariffs on calls that they receive.

The charging scheme that is used does affect the rate of calling. Originating a call and receiving a call are imperfect substitutes that may nevertheless have high cross-elasticity. Technology may enable a party who wishes to have a conversation to signal the intended correspondent to call him. For example, pagers can serve to notify cellular users of callers attempting to reach them; the cellular user can then return only high-value calls from the mobile phone. Country-direct international service enables caller A to connect to subscriber B in a second country and be charged at the B-to-A tariff.

An integrated supplier bears all of the costs of the A to B call and receives all revenues. His objective is to maximizes the combined profits of the two sub-networks. A menu of pricing options—including caller billing, toll-free calling, calling cards, and optional calling plans—has developed.

Now consider the case of two separate network operators. In the special case of balanced traffic (A to B calls equal B to A calls) and equal costs in each network, all payment schemes recover the same revenues. In general, however, traffic flows and costs will differ. Interconnection pricing then becomes a positive-sum game with unequal payoffs:

- interconnection increases the total surplus from use of each network (the network externality)

- subscribers to the smaller network enjoy a larger per-subscriber increase in surplus from having access expanded to a larger number of subscribers.

- interconnection is entirely complementary to the autarchic network's demand.

1.2.2 Externalities

The interconnection of networks N1 and N2 provides the subscribers of each network with a valuable service—the opportunity to communicate with members of the second network.

It gives rise to two externalities. The call externality is the benefit of a call to the party that does not have to pay for the call (usually the party being called). The network externality, which results from adding new consumers to the reach of the original network, is the aggregate benefit that other subscribers enjoy by being able to call, or to be called by, the new subscribers.

The network externality can be measured by the sums of the consumer surplus changes of other customers who call and who are called by the new customers. If the number of calls between interconnected networks is known, the aggregate increase in willingness to pay of subscribers in N1 measures the value of access to N2. In addition, there may be some increase in surplus for uncertain calls to the interconnected network. Access to a broader market may stimulate new services and promote toll-free calling.

The call externality is often assumed to be internalized by correspondents. When A and B have an established relationship they may relatively easily arrange to share the costs of calling. For certain categories of calls the called party values the call more than the caller. The vigorous growth of the 800 number services demonstrates that in many other cases the call externality can be appropriated by the second party. Extending toll-free calling across an interconnection can be a significant part of the value of interconnection.

1.2.3 Overlapping networks

Suppose, now, that the N1 and N2 networks serve the same region and compete for subscribers. (Figure 2). If we set aside economies of scale due to density, cost conditions will be similar to the case of disjoint networks considered above. Now interconnection increases the total surplus in the market that is realized from the network externality, but only for subscribers in the non-overlaping areas. However, there may be additional surplus gains due to price and service competition. Interconnection is at least a partial substitute for the demand faced by an autarchic network.

1.3 Functional view of interconnection

A telecommunications operator (TO) may be considered a self-contained network in which its users enjoy full connection to each each other for access to a basic set of network services. The design and operation of one network requires solving a list of key issues, including network management, user numbering, traffic routing, signaling, transmission, pricing of services, and so on.

Interconnection of two TOs requires either a common solution to each issue, or else the adoption of standards enabling each network to translate the practice of the other. Interconnections among TOs, service providers, and providers of network components will raise many of the same issues. Figure 3 shows a schematic representation of these issues.

Successful interoperation of networks requires not only the physical interconnection of transmission links, examined in the context of physical interconnection later, but successful meshing of engineering, operational, maintenance, routing, and addressing systems and procedures. Functional interconnection as posing issues at each of the following levels:

- Management: Policies for interconnection in operations, administration, maintenance, and provisioning are needed to ensure rapid diagnosis and recovery from network faults, and efficient and expeditious ordering for facilities or services which require multiple providers. Network management (fault diagnosis, fault recovery, security, accounting, and performance management) requires common protocols and procedures among TOs, mutual provision of data to a management information database, and cooperation in supporting loopback services, or other diagnostic procedures.

- Service quality: Network blocking due to insufficient capacity to overcome traffic congestion in one network reduces real-time service quality and may induce addition congestion from repeated call attempts. Interconnected networks may require coordinated demand and capacity forecasting and cooperative real-time management controls to limit congestion when it develops.

- Location: Policies may be required to ensure appropriate levels of cooperation in siting of network facilities. For example, decisions on the number and locations of access tandems in local exchange service areas can have significant impacts on the efficiencies of interconnection with the points of presence (POPs) of multiple long-distance carriers. Physical collocation of network terminating equipment at a TO facility, or virtual collocation at a nearby point without extra charge, may be necessary to ensure that competing carriers have access to TO resources on equal terms.

- Numbering: The telephone numbering space is a scarce resource that requires procedures for partitioning and management so that users can obtain telephone numbers in an interconnected public network. Currently, the assignment of numbers is typically partitioned among TOs. As customers are expected to demand local number portability in the future, number assignment must be coordinated among operators. Proposals for new numbering assignment principles which are not geographically based will require new numbering administration policies which are operator neutral.

- Routing: Routing functions can be accomplished on a hop-by-hop basis using local information maintained at each switching node, or can be optimized more globally and in real time through the use of intelligent network functionality, with routing logic executed at service control points. In the former case, each operator of a switch is responsible for updating and maintaining routing information for its switches. Policies governing these routing tables may have to be established to ensure, for example, customer choice of alternative carriers, or fair allocation of unspecified traffic among alternative transport providers. With centralized routing information, more complex routing rules become possible, including those which vary carrier choice by time of day, traffic levels or cost. Centralized routing servers are also likely to be required to implement provider-independent numbering, e.g. for local and toll-free number portability.

- Signaling: With the rise of common channel signaling, the signaling systems of service providers must be interconnected as well as the transmission systems. Alternatively, one might imagine one entity operating a signaling network that is used by the switches and transmission facilities of another entity. The US cellular industry provides an example in which multiple cellular operators are contracting for a third party to operate a common channel signaling network for their use.

1.4 Physical relationships view of interconnection

The physical components of the voice telephone network consist of discrete elements (e.g. switch, database, peripheral computer) that are arranged into a network architecture and connected by physical facilities for voice transport and signaling. In Figure 4 these components are represented schematically, with points of interface between components shown as circles.

Two networks are physically interconnected at one or more interface points where voice circuits and signaling paths of each network connect to each other. At the interface one or both networks may deploy electronic and software capabilities to screen the electrical signals and control messages received from the other network in order to ensure conformity with its equipment and operating standards.

Service providers, users, and other entities may, in addition, interconnect physical components (e.g., a private switch, a transport link, or an intelligent peripheral) with the TO network.

The conventional network architecture is depicted in simplified schematic form in Figure 5. Beginning at the subscriber, the first element of the network is terminal equipment—a telephone handset or PABX - which is connected by wiring inside the user's premises to the TO network termination point (A). From that point the subscriber is connected by the distribution network to the TO local exchange.

The distribution network (outside plant) consists of large feeder cables connecting the local exchange to a serving area interface located within several hundred feet of the subscriber, and then smaller distribution cables to individual subscribers. Fiber optic technology is replacing many feeder cables, and subscribers with high communications volumes, such as larger business offices, may have direct fiber connection to the carrier's office. Interconnection in the distribution system, not generally available, could be demanded by a wireless network.

For example, a personal communications service radio microcell could serve the area of subscribers now connected through the TO distribution plant and interconnect to the network at a serving area interface (B). At the local exchange the subscriber line terminates at the subscriber main distribution frame (C). This "line side" interface to the local switch is a point of interconnection in US and UK networks for a competing local transport carrier who may connect a subscriber to the local switched telephone system.

Within the local exchange, the subscriber's line at the main distribution frame is connected to the local switch. In the case of a leased line, the connection is to a cross-connect device. On the "trunk side" of the local exchange, interoffice trunk facilities (copper cable, microwave radio, or fiber optic cable which carry aggregated traffic or high-volume data) terminate at the trunk main distributing frame, from which they are connected to the switch, or to a cross- connect device.

The trunk main distribution frame (D) provides interconnection for long distance carriers, information service providers, and competitive access providers. In the TO network, interoffice trunks connect a local exchange to a tandem and switch (as shown), to another local exchange, or to the point-of-presence of another TO network. At a tandem switch, trunk terminations at the distribution frames (E, F) are additional interconnection points.

A competitive access provider could serve high-volume business subscribers with a fiber-optic cable that it interconnects to the network. To do so it would collocate its own termination equipment in the TO local exchange and then connect that equipment to the subscriber main distribution frame. A cellular radio network connects its antennas located at distributed cell cites to a mobile telephone switching office. That office is then interconnected with the TO network by trunks that terminate at either the tandem office (E) or the local exchange (D).

The physical interconnect model serves to identify the architectural points of interface at which networks and service providers can potentially interconnect. A realistic view of the variety of interconnection arrangements that are now being requested in the local TO network may be gained from Figure 6, derived from a detailed study of the issues by the US Information Industry Liason Committee (IILC). In the figure the dashed lines indicate traditional TO service to the end user, while solid lines depict various types of interconnection for other entities.

Some 13 types of demands for access to the local switch may be grouped into five major categories, according to the network interconnection point at the TO's wire center.

1. Customer Premise Equipment Interconnection: The end user connects "inside wiring" and terminal equipment to the network, or the user connects a private switch (PABX) to carrier-supplied trunk-termination equipment at the user's premises.

2. Outside Plant Interconnection, at the Serving Area Interface: A distributor or alternative local network (1) provides a dedicated connection to the end-user, e.g. by cable or radio, and transports its traffic to the TO wire center by TO feeder cable.

3. Subscriber Main Distribution Frame Interconnection: A competitive access provider supplies all feeder and distribution connections to the end user and interconnects copper or fiber facilities to line side of the TO switch (2). In other instances the provider uses the TO's feeder and distribution cables for its own services, but needs access to the distribution frame to connect to the cables at the switch (3, 15).

4. Trunk Main Distribution Frame Interconnection: The non-TO provides transport from the trunk side of the TO switch to some other entity, using digital copper or fiber facilities. The transport entity may itself provide competitive local switching and distribution (4), PBX switching at the end-user's premises (5), transport and switching for radio-based access services (cellular, paging, etc.) (6), or transport to a long distance carrier's point of presence (POP) (7), or transport to another TO local exchange or tandem (8,9). An alternative local transport service provider, with tandem function, could also interconnect its switch, including addressing capability, to the TO switch (10).

5. Digital Cross-Connect System Interconnection: The non-TO supplies leased line facilities to end-users and interconnects with the TO to perform remote network reconfiguration at one of several physical connection points in the TO switch (the fiber distribution frame, digital cross-connect, main distribution frame) (12). Alternatively, the provider uses the TO's controller capabilities to service the leased line facilities that it purchases from the TO (13).

The physical interconnect model highlights fundamental interconnection issues, including:

- the need for standardized physical and electrical interfaces in the transport and switching facilities.

- the dimensions of comparable, if not fully equal, access to physical network components.

- numerous possibilities to unbundle the physical network into modular components that may be supplied by competing entities.

1.6 Logical relationships view of interconnection

A different perspective on network interconnection flows from the view of networks and service providers as suppliers of functions, arranged in a logical architecture, that control communications pathways and access to information. In Figure 7 these logical functions are schematically related to the physical telecommunications resources, with points of logical interface shown in circles.

Two networks are logically interconnected at one or more interface points when one network's logical resources provide control or information to the other. For example, one network may provide a second network with access to its customer information for directory inquiries and call routing. Service suppliers and end users may also logically interconnect their own information and management resources at these interfaces, for example, to route terminating green-number calls over an interconnected private network.

Control of calls and access to essential data within the TO network depends on the network signaling system. The conversion by TOs from in-band signaling to the common-channel signaling system (SS7) will make possible standardized interfaces for logical or programmatic (in contrast to physical) interconnection with the TO network. For example, an independent information services provider could provide a database that would interconnect through a service control point. Another enhanced service provider could access a TO database, using defined SS7 messages, independently of a telephone call.

These forms of logical interconnection raise policy issues similar to those of physical interconnection, including standardization of software protocols, equal access by TOs and other service providers, and unbundling and pricing of access to logical resources. The translation, or mapping, of these demands for various types of logical interconnection onto the TO network has not yet been worked out.

1.7 Key access relationships

The fundamental relationship between the incumbent network operator and a potential entrant network operator is interconnection—the ability to deliver traffic originating in one network for termination in the second network. A dominant incumbent who supplies the vast majority of access lines to end users controls a bottleneck resource that constitutes essential facilities for an entrant seeking to terminate traffic to those users. If the incumbent denies interconnection he can effectively foreclose the entrant from reaching most of the market. Less extreme measures that allow interconnection, but on conditions that are less favorable than those the incumbent enjoys, effectively raise the entrant's costs. The most significant conditions are the locations at which interconnection is permitted and the charges for establishing points of interconnection and for conveyance of interconnected traffic.

The quality of access affects supplier relationships in several ways:

In its initial stages of expanding its own network an entrant generally requires the right to resell the incumbent's services in order to reach areas of the market in which it has not yet installed its own network facilities. For long-haul transmission services, this can generally be accomplished by leasing high-capacity facilities from the incumbent. For access to end users, it requires interconnection to the incumbent's network to terminate traffic for delivery at distant local exchanges.

The incumbent telephone networks have not been designed for a multiple-supplier environment. Providing the subscriber to the incumbent network with access to competing network carriers requires modification of local-exchange switches, extension of routing tables and software to accomodate carrier codes, and a method to allow the subscriber to designate the carrier of choice.

Fully equal access arrangements would present the subscriber with a symmetric arrangement for all carriers. One method is to prefix every call with a carrier code that is used by the local exchange to route the call to the selected carrier. As a requirement for every call, this represents a significant user cost to subscribers who intend to continue using the incumbent network as their carrier. Preselection or presubscription allows default selection of carrier for each subscriber line. In the US the presubscription form of equal access was implemented by upgrading all local tandem switches.

Equal access arrangements incur significant costs for the incumbent network and are only possible where stored-program control switches are in place. Their effect on competition is not unambiguous. In the US market structure, with separation of local exchange and interexchange carriers, equal access eliminates important quality differences between interexchange carriers and intensifies price and feature competition. In the UK and Australia, for example, integrated local/interexchange networks compete for retail customers. In this setting, differences in the quality of access to end users can stimulate a competing network to develop its own local access facilities, leading to direct competition in the supply of access service.

Equality of access extends beyond the physical conditions of interconnection to include provisioning and maintenance as well as access to customer information of value in marketing other services. Without regulatory oversight the incumbent operator is in a position to provide systematically higher quality to its own customers that to a competing network operator.

Successful interconnection also requires a number of additional assured relationships for related conditions and services between the network operators. Technical interfaces must adhere to agreed-upon standards. Network operation must achieve a reliable level of quality and be regularly maintained. Signaling messages, as well as traffic, must flow between the networks without impediment. For many value-added services, database queries must be processed and information updated between the operators.

1.8 Related issues

Numbering of telephone access lines is traditionally administered by the incumbent network. With competitive access suppliers, issues of fair administration have arisen. More difficult is number "portability," which would allow a subscriber to retain his telephone number when switching to a different access network supplier.

Under regulatory requirement the US networks have recently made toll-free 800 numbers fully portable, so that, for example, a business that nationally advertises its 800 number can change network carrier and retain the same number. The required database system became possible when both local exchange and interexchange carriers had completed conversion to the common-channel signaling system for all interexchange traffic. When each 800-number call is dialed, a regional control point looks up the associated telephone number in the database, and the call is then completed to that number.

Full portability of local numbers would require a database capability available to each local exchange and number translation for every number dialed. However, operational issues and the extensive investment that would be required make full number portability unlikely for at least some time.

2.0 COST STRUCTURE OF INTERCONNECTION

The costs of providing the telecommunications network can be split into two basic categories: non-traffic sensitive costs—which consist of overhead costs and costs that vary with the number of subscribers—and traffic-sensitive costs—those which vary with either the peak load or the total volume of calling.

2.1 Wireline incumbent network

The non-traffic sensitive components of the network are primarily the line that connects the subscriber's telephone instrument to the local exchange and the equipment at the local switch that terminates that line. These so-called "local loop" facilities are reserved for the subscriber whatever his level of traffic. In most telephone networks subscriber rental charges cover only a portion of the the non-traffic sensitive costs. The balance is recovered from rates for other services, interconnection, and in some cases higher charges for business rentals.

The traffic-sensitive components of the network are sized to have capacity sufficient to carry the maximum traffic that is ordinarily present in each section of the network. The rate of use of the network varies considerably over time, and the profile of daily traffic follows a pattern such as the one in Figure 8. A very high proportion of the costs of carrying the traffic over the network are caused by provision of switches and transmission links large enough to accommodate the peak traffic rate. If busy-hour traffic increases, the network facilities must be expanded. However, if traffic increases in other hours, there is unused capacity available. As a result, the traffic-sensitive costs of the network are almost entirely determined by the peak traffic loads. In addition, there are some much smaller costs that vary directly with the number of calls, so that except during the busy hour the cost of carrying traffic is very low.

Calls that originate in one network and terminate in the other incur additional costs of interconnection. The point of interconnection used for a particular call depends on the location of the originating and terminating switches. Calls are usually routed so that they will be delivered to the second network at the point of interconnection close to the destination subscriber. Thus, in Figure 9 a call from a subscriber A in Network 1 to Subscriber B in Network 2 could enter network 2 at POI3 and use traffic-sensitive components that consist of tandem switching, transmission to the local exchange, and local exchange switching to deliver a call to subscriber B. In addition, the call would use the non-traffic sensitive subscriber line and the equipment reserved for that line in the local exchange.

The interconnecting traffic will only increase the capacity costs of Network 1 if the additional traffic increases that network's maximum level of traffic beyond the level already carried during the busy hour.

2.1.1 The concept of the busy hour

To achieve a key aspect of the quality of communications service—a low rate of congestion—the capacity of the network routes serving each origin-destination pair of access points must be sufficient to accommodate nearly the maximum simultaneous rate of traffic on that route. The traffic flow varies in regular patterns. Over a 24-hour period traffic on a given route typically reaches its peak at a consistent hour. On routes with predominantly business traffic the busy hour is most often in the late morning; on residential routes it may occur in late afternoon or early evening. Weekend and holiday traffic is generally lower in business areas. Traffic patterns also vary by location as determined by the predominant local activities; thus, city centers and suburban or rural areas often have different busy hours.

In addition to systematic traffic patterns, demand is influenced by many random factors, including unforeseen economic events and singular occurrences such as national disasters and charity phone-in programs.

To determine actual traffic loads on a network and to plan for capacity expansion, traffic engineers deploy automatic recording equipment during the busiest season of the year for the section of the network to be observed. Continuous traffic observations are collected for a two to four-week period. The data are averaged by day of the week to remove random variations and the busy hour for the measured network component is then determined. Separate measurements are made for representative network components—switches and transport links—in each segment of the overall network. In addition, traffic flows at major network notes are continuously monitored. If congestion occurs, operating personnel may be able to redirect some traffic over alternate routes or restrict the number of calls allowed to enter the network from a particular area.

2.1.2 Network cost patterns

The costs of the network are composed of several major factors—sufficient capacity to convey traffic between origination and destination notes, capacity to obtain access to those nodes, capacity to provide optional telecommunications features associated with calling, and staffing for operator-related services. These costs are driven by several factors:

- capacity costs of conveyance are determined by the number of calls initiated (setup) and the duration of conversations at the busy hour of each network component. The pattern of capacity costs follows the pattern of busy-hour traffic.

- variable costs, for call accounting and some miscellaneous activities, are determined by the total volume of traffic.

- access costs, which include the local loop, subscriber line database information, and billing activity, are determined by the number of subscribers, their distance from the local switch, and the type of construction used for loop facilities. These costs do not vary with the volume of traffic.

- features costs, for providing caller line identification, call forwarding, and similar features, are determined by the number of users who subscribe for those optional features and do not vary with traffic volume.

- operator and directory enquiry costs are determined by the number of calls that require operator intervention for billing or person-specific termination.

Traffic-sensitive costs are costs that increase with greater volumes of either busy-hour traffic or total traffic. The costs of traffic conveyance are traffic-sensitive and almost entirely determined by the levels of busy-hour traffic in each network segment. Operator services and database-related services are traffic-sensitive costs that are driven as well by total traffic volumes.

Traffic-insensitive costs do not increase with either busy-hour traffic or total traffic. Due to the indivisibility (lumpiness) of facilities, there is sufficient capacity at the minimum scale of traffic-insensitive equipment to accommodate large increases in traffic. These costs are associated with the telephone instrument and premise wiring, the local loop, and the line-terminating equipment in the local switch.

Although aggregate network traffic statistics are useful for obtaining a macro view of network use and changes in patterns over time, but they do not provide an accurate guide to cost changes or investment planning. Costs are incurred at the level of individual network components. Planning for network expansion and modernization is based on the analysis of existing capacity, current busy-hour traffic, and expected growth in busy-hour traffic for each switch and for trunk groups between each pair of nodes. Annual network investment is the sum of a large number of individual expansion projects.

Network planning recognizes that, in addition to regular daily traffic patterns, telephone calling is subject to a host of random factors as well. To reduce the chance of excessive congestion when an unexpectedly large volume of traffic occurs, designers include a margin of spare capacity in excess of the predicted busy-hour traffic. Margins are calculated for individual network links and switches, taking into account the statistical laws of traffic aggregation and, where available, the possibility of alternate network routes.

2.1 3 Effect of interconnection on network costs

Interconnection may increase the total traffic flows on the incumbent network, change its distribution over time and among regions and also reduce flows in some segments of the network. To the extent the entrant's traffic increases the total traffic on the incumbent's network at busy hours, it imposes new costs. The incumbent network must be expanded to avoid congestion and degradation of service. Provided that the magnitude of the additional demand is known in advance, the amount of new capacity required is equal to the increment in demand; no margin is needed for demand variation. The additional network costs are the incremental costs of expanding capacity.

To the extent the entrant's traffic displaces the incumbent's traffic, it substitutes for previously carried traffic. While originating on a different network, this traffic requires the same busy-hour capacity as the traffic it replaces and does not impose new network costs for the incumbent, who has previously incurred the costs of providing busy-hour capacity sufficient to meet the pre-entry level of demand.

2.2 Wireless network.

The cost structures of cellular and personal communications systems networks differ significantly from the cost structure for fixed wireline networks. Radio spectrum is a scarce resource to these networks, limited either by government allocation or the cost of bidding spectrum away from other uses. The spectrum is shared by radio trunking technologies and spatial division of the serving area, with a pattern of radio cells designed to reuse frequency bands in non-adjacent cells. The maximum capacity of cellular networks can be increased by reducing cell sizes (increasing the costs of antenna sites and backhaul trunk facilities) or adding frequencies. End-user access requires occupancy of a frequency band in proportion to conversation time. Consequently, the structure of access costs in a wireless network is strongly sensitive to peak usage.

3.0 ALTERNATIVES FOR PRICING INTERCONNECTION

Interconnection with a second network can be viewed as buying two related services: access to all of its subscribers, and use of the second network as traffic conditions demand. Thus the demand value of one network operator for interconnection with another has an option component and a use component. Confronted with this demand structure, a supplier of interconnection would be likely to develop a multi-part pricing structure.

Viewed from the supply side, the costs of providing interconnection fall into two broad categories—the direct costs of supplying interconnection service to a second network operator, and the effect that interconnect competition will have on the costs of the public service obligations of the incumbent operator.

The incumbent network operator has developed the access network in monopoly setting. Costs and rates are not well aligned. Rate averaging across regions with disparate costs. Cross-subsidies and price discrimination are typical between long-distance and local access, and business and residential access. Political resistance rates has prevented a rebalancing of rates to align them at least roughly with costs.

We consider several interconnection pricing approaches that have been advanced.

3.1 Incumbent opportunity cost pricing

Assume that the market for retail services is contestible by a second network entrant, who nevertheless requires an input—interconnection—from the incumbent network. If we further assume that the incumbent network is an efficient, multi-product monopoly, its retail pricing structure will embody Ramsey pricing. In general terms, inelastic services will have high price-to-marginal cost margins, elastic services will enjoy small margins, and access services may have varying margins reflecting network externalities and patterns of demand elasticities. Non-linear pricing, in the form of two-part tariffs, volume discounts, and service bundling, will develop market segmentation to a high degree.

It is efficient to have retail service produced by the lowest cost supplier, either the entrant or incumbent. In these circumstances this outcome is achieved by pricing, interconnection, the essential intermediate component in the retail service, at the opportunity cost of the incumbent. Here, the opportunity cost is equal to the competitive retail price minus the incumbent's incremental cost of the interconnection component. (Baumol, forthcoming).

Implementation of opportunity cost pricing by the incumbent would incorporate retail tariffs into the calculation of interconnection prices. It would also require information about the cross-elastic effects on the entrant's services. Note that this pricing scheme would discriminate among interconnecting different networks that used the same components and imposed the same costs, according to their differing effects on the incumbent's retail services.

A fundamental shortcoming of opportunity-cost pricing is that it is founded on the assumption of efficient pricing by the incumbent network. The opportunity costs of this approach do not distinguish between the opportunity costs that result from efficient pricing to reflect network externalities and differential demand elasticities for a supplier with some economies of scale or scope, and the revenue opportunities from monopoly profits and inefficiencies of monopoly supply that the incumbent network operator would incur.

3.2 Tying pricing to retail tariffs

In the UK competition was introduced in 1983 with an explicit policy of duopoly. The entrant national network operator, Mercury, was granted a license with assurance of no additional entrants for a seven-year period.

In 1985 the regulator, Oftel, issued a determination governing interconnection of BT and Mercury. Points of interconnection are permitted at the line side of local exchange end offices (3L) and the trunk (junction) side of tandem offices (3J). Mercury pays for direct costs fo connection, or 50% of fully allocated costs if BT has excess capacity at the connection point.

Charges for interconnected calls have followed the structure of BT's retail tariffs. Charges have the three time-of-day periods (peak, standard, and cheap) and three distance bands (local, short national, and long national). The local interconnection rate distinguishes the two types of interconnection points (line side, trunk side).

One effect of the UK interconnection pricing structure is to direct the entrant's competitive efforts toward customers making peak-period long-distance calls. The interconnection price represents a floor on Mercury's costs in each rate period and distance band. For local calls, BT's retail price in the cheap period has been equal to the interconnect charge. Margins on local calls are 35-40 percent in the peak hours, rising to some 50-75 percent in peak hours for short and long national calls. Over time, BT has rebalanced retail rates to the extent permitted by the overall price cap, reducing Mercury's potential margins.

segment	cheap rate	standard rate	peak rate
Local 3L	1.023	2.352	2.659
Local 3J	1.023	2.046	2.352
short national			
A	2.250	4.5	6.136
B	1.18	3.17	4.705
long national			
A	2.455	5.114	6.648
B	1.?	3.58	4.9?

3.3 Incremental cost pricing

Australia has established three categories of charges for interconnection:

(1) establishing physical interconnection: costs incurred in transacting interconnecting and achieving equal access

(2) carriage of traffic on the other carrier's interconnected network

(3) supplementary and ancillary services, not concerned with call routing, i.e. monitoring, recording, fault reporting, billing.

The charges are calculated based on long-run incremental costs, including fixed capital costs of relevant assets and directly attributable recurring operating and maintenance costs. They are levied on a network-component basis, with charges differentiated by area (central business district, motropolitan, and country). The nature of costs as determined by capacity requirements is weakly reflected in time-of-day charges based on a 10- or 14-hour weekday peak period.

In addition to charges for incremental costs of interconnection, the entrant contributes to the Universal Service Fund through charges based on minutes of use.

3.4 Capacity cost pricing

In the UK Mercury is seeking to have traffic-sensitive charges for interconnection based on BT's fully-allocated costs of capacity at the busy hour. Under its proposal, Mercury would forecast quarterly traffic profiles for each of BT's network segments and pay in advance the fully-allocated capacity cost of its forecast busy-hour load for each network segment. Traffic flows that exceeded the forecast would be charged at a penalty rate, perhaps equal to the retail tariff. Since there would be no refunds for delivering less traffic than forecast, this pricing arrangement is similar to a guaranteed purchase of capacity, on a take-or-pay basis.

In addition to busy-hour charges, Mercury would pay a per-minute Access Deficit Charge (ADC) calculated from BT's fully-allocated cost of access less subscriber rental charges. The ADC is allocated to minutes of traffic according to the retail rates prevailing in the peak, standard, and off-peak charging periods and distance bands. 3.5 Uniform per-minute pricing

In the US, the Modified Final Judgement required that charges for interconnection between end offices and interexchange carriers be equal, per unit of traffic. The equal-charge provision consciously provided assistance to new interexchange entrants, and removed any basis of charging for interconnection on the basis of cost orientation. Thus, a new entrant and AT&T paid the same rate for each minute of traffic, irrespective of the costs of the components of a local exchange carrier used, the facilities actually operated, and the economies of scale they enjoyed.

Even at comparable volumes, per-minute pricing is only weakly related to network costs, since nearly all of the incremental costs of the components needed are due to busy-hour capacity requirements.

Following the expiration of the MFJ equal-charge provision, the FCC has moved to base access charges more directly on cost components. For special access services, there are three rate elements:

(1) a flat-rate, non-recurring connection charge

(2) a recurring channel termination charge for each connection between the customer and between the central office, and the central office and the interexchange carrier's point of presence

(3) a distance-sensitive per-minute charge for any required links between central offices.

For switched access service, the FCC ordered an interim rate structure consisting of four elements. Local exchange carriers have generally proposed tariffs with the following features:

(1) entrance facilities charges—a flat rate without volume discounts

(2) direct-trunked transport charges—flat monthly rate for termination, plus distance-sensitive rate.

(3) tandem-switched transport charges—per-minute and distance-sensitive rate

(4) a residual interconnection charge—per minute, intended to recover the remaining costs of switched access transport,

Uniform per-minute interconnection charges are a crude form of specific commodity tax. By unbundling US access charges, rates will now recover some capacity costs by direct capacity-related charges. However, other rate elements persist in spreading busy-hour capacity costs across usage at all hous and recovering access deficits by rates that bear no relationship to value of the service demanded.

3.6 Wholesale vs. retail pricing

At the wholesale level interconnection sales are significant to both the incumbent and the entrant network operators. Interconnection is a major cost of the entrant's business, so that he is acutely price sensitive. Costs of negotiation and evaluation of detailed pricing structures are small relative to cost reductions that may be achieved. Interconnection charges translate directly into cost of service for the entrant's network.

Basing interconnection charges on costs encourages economic efficiency:

- in the final pricing of the entrant's services

- in the entrant's choice of production activities

- in facing the entrant and the incumbent with the same marginal costs of expanding service for the network components used by both.

End users consume much smaller quantities of telephone services. Compared with the potential cost savings that might be achieved, the costs of searching out the lowest cost tariff loom large. Suppliers compete to differentiate their service offerings along additional dimensions — quality and matching the tastes of each market segment. This competition provides increased welfare not captured in the assumption of a single homogeneous product assumed in the pure theory of peak-load pricing.

Leased lines are an example of alternative end-user pricing that is consistent with cost-based network-service pricing. Consider a customer who purchases the capacity to connect two business locations for a fixed monthly rental based on the full amortized cost of the facility. His use of this very small private network and allocation of its costs to individual callers involves a pricing decision for end users. If the line's busy hour traffic is high, congestion will create an opportunity cost during the busy hour. The customer may choose to implement an internal peak-load pricing scheme for calls. Or, he may find an alternative a more effective way to maximize the benefit of this resource, by establishing priority of access, blocking some users from access at peak hours, or allowing service quality to vary with traffic intensity.

At the retail level tariffs are considerably simplified from the optimal marginal-cost and peak-load tariffs of economic theory. For example, end-user time-of-day tariffs are limited to not more than three different price levels per day. Such tariffs necessarily involve a considerable averaging of costs and do not achieve the efficiency of theoretical peak-load pricing (Park and Mitchell, 1986).

4.0 Interconnection Policy

In local telephone markets interconnection lies at the heart of a policy for competition. In these markets access to components of the local telephone network operator's facilities and services is essential to successfully launching new entrants.

An effective policy will have two central thrusts: unbundling of the network to permit interconnecting operators to purchase only the inputs they require, and pricing of interconnection to encourage market efficiency.

We have noted the considerable variety of operators who seek physical interconnection with the incumbent local network. A survey of information service providers would undoubtedly show a similar diversity in the demand for logical interconnection. The Open Network Architecture proceedings have begun the unbundling process in the US, but achieving an effective degree of modularity will probably first require widespread diffusion of an "intelligent network architecture.

Efficient interconnection pricing is hampered by the overhang of older policies to support low access rates and political reluctance to rebalance rates more generally. Absent rebalancing, interconnecting operators are seen as an essential source of revenue to fulfill access and universal service objectives. Regulatory authorities and the courts have variously promoted interconnection rates that favor entrants. In the UK, payment of the access deficit charge can be waived until the entrant gains a 10 percent share of the market. In the US access rates based on average per-minute traffic measures have only slowly be realigned to orient them toward costs.

Market entrants bring the prospect of considerable scope for innovation in two dimensions—assembling telecommunications components into new services not supplied by the incumbent, and matching pricing of services to the preferences of final consumers. From this perspective, basing interconnection prices on the incumbent's retail price structure provides little room for such competitive responses, and prices based on the incumbent's opportunity costs are also highly restrictive. The interconnection unbundling and pricing schemes to be preferred are those that offer the entrant a wide degree of flexibility to innovate.

REFERENCES

W. Baumol, "Deregulation and Residual Regulation of Local Telephone Service", American Enterprise Institute, forthcoming.

M. E. Beesley and B. Laidlaw, "The British Telecom/Mercury Interconnect Determination," in, M. E. Beesley, *Privatization, Regulation and Deregulation*, London, 1992.

G. Brock, "Interconnection Conditions, Access Charges, and Universal Service", TPRC, October, 1993.

M. Cave, "Interconnection, Separate Accounting and the Development of Competition in UK Telecommunications", Institute of Economic Affairs, Department of Economics, Brunel University, 12 November 1993.

Bridger M. Mitchell and Ingo Vogelsang, *Telecommunications Pricing: Theory and Practice*, Cambridge University Press, Cambridge, 1991.

M. Mueller, "Comments on Interconnection for European Telecommunications", 21st TPRC, October 1993.

W. Neu and K-H. Neumann, "Interconnection Agreements in Telecommunications", Discussion Paper Nr. 106, WIK, Bad Honnef, April 1993.

R. E. Park and Bridger M. Mitchell, "Optimal Peak-Load Pricing for Local Telephone Calls," RAND R-3404-1-RC, March 1987.

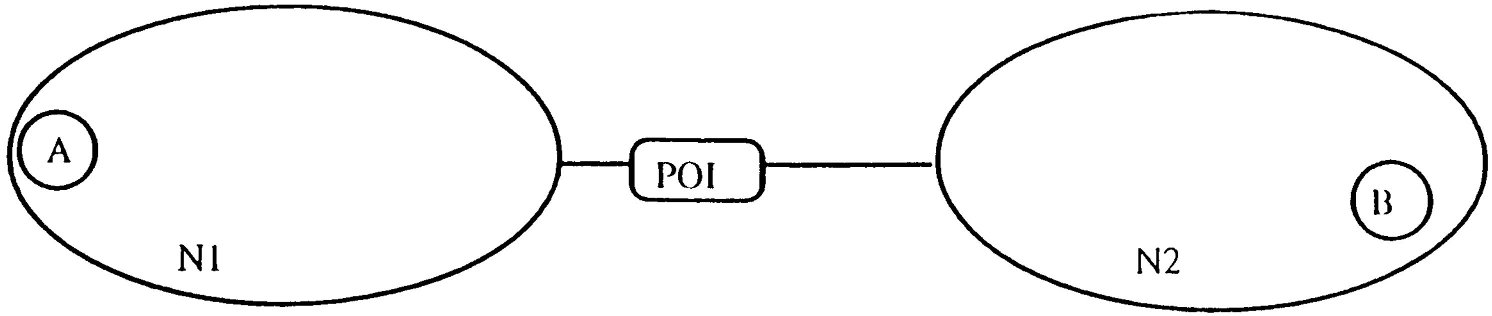


Figure 1

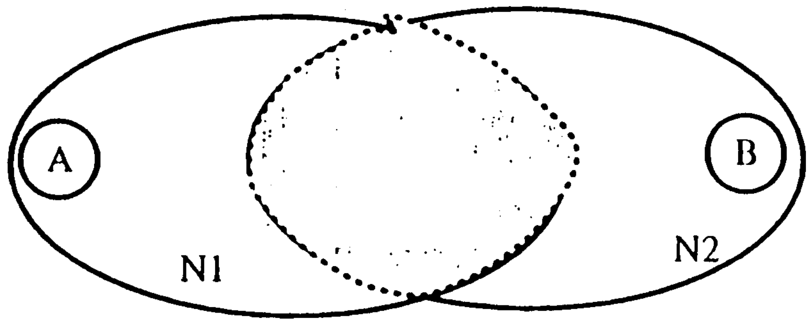


Figure 2

Figure 3 Functional interconnect model

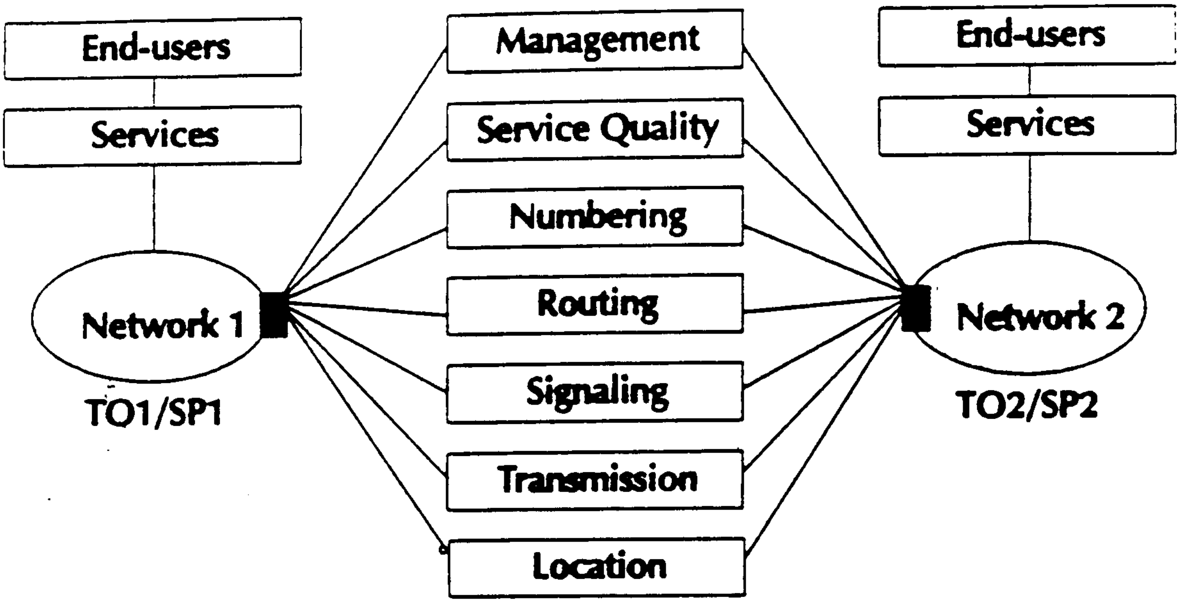


Figure 4 Physical interconnect model

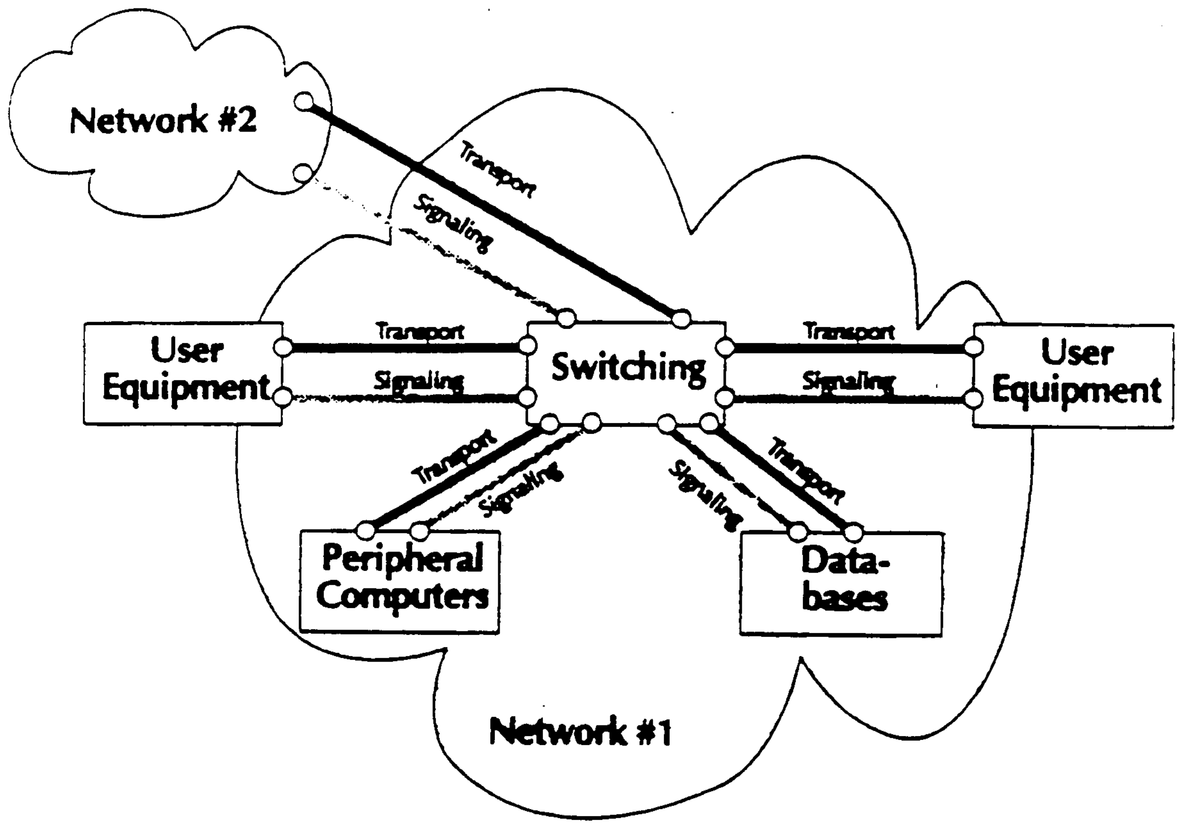


Figure 5 Architecture of a telephone network

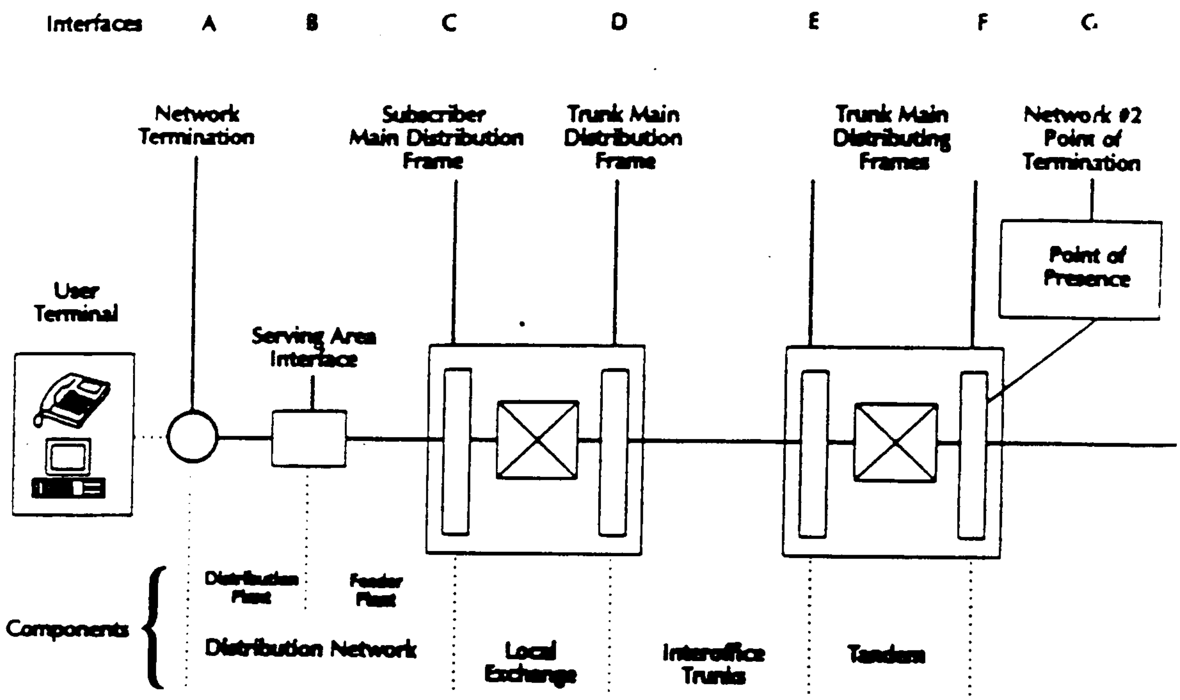


Figure 6 Interconnection arrangements requested in local TO network

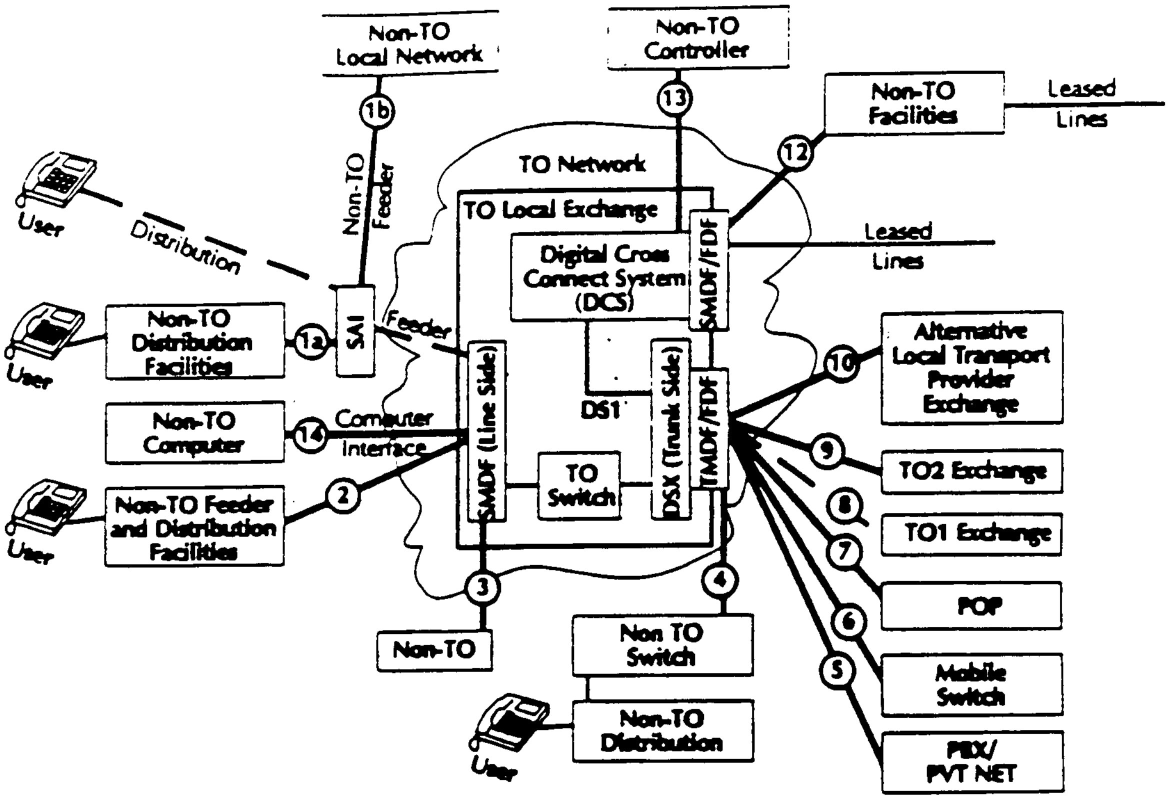
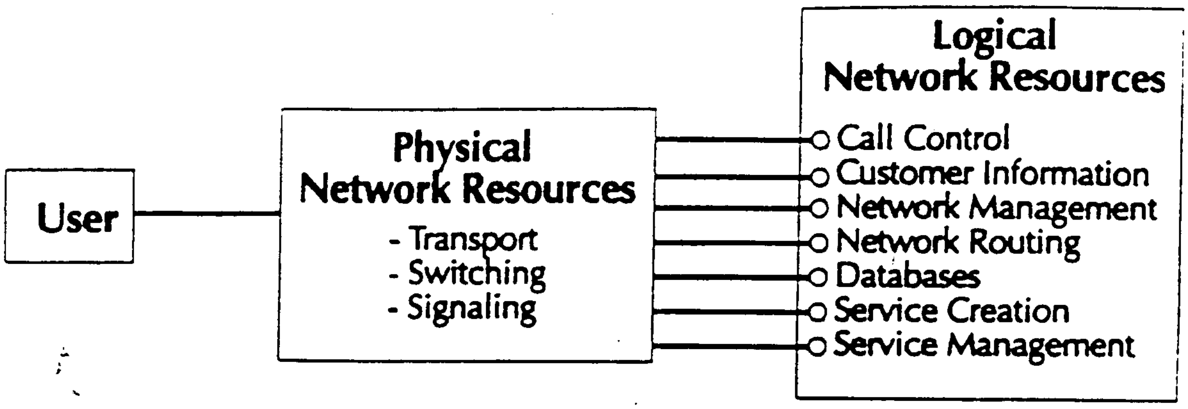


Figure 7 Logical interconnect model



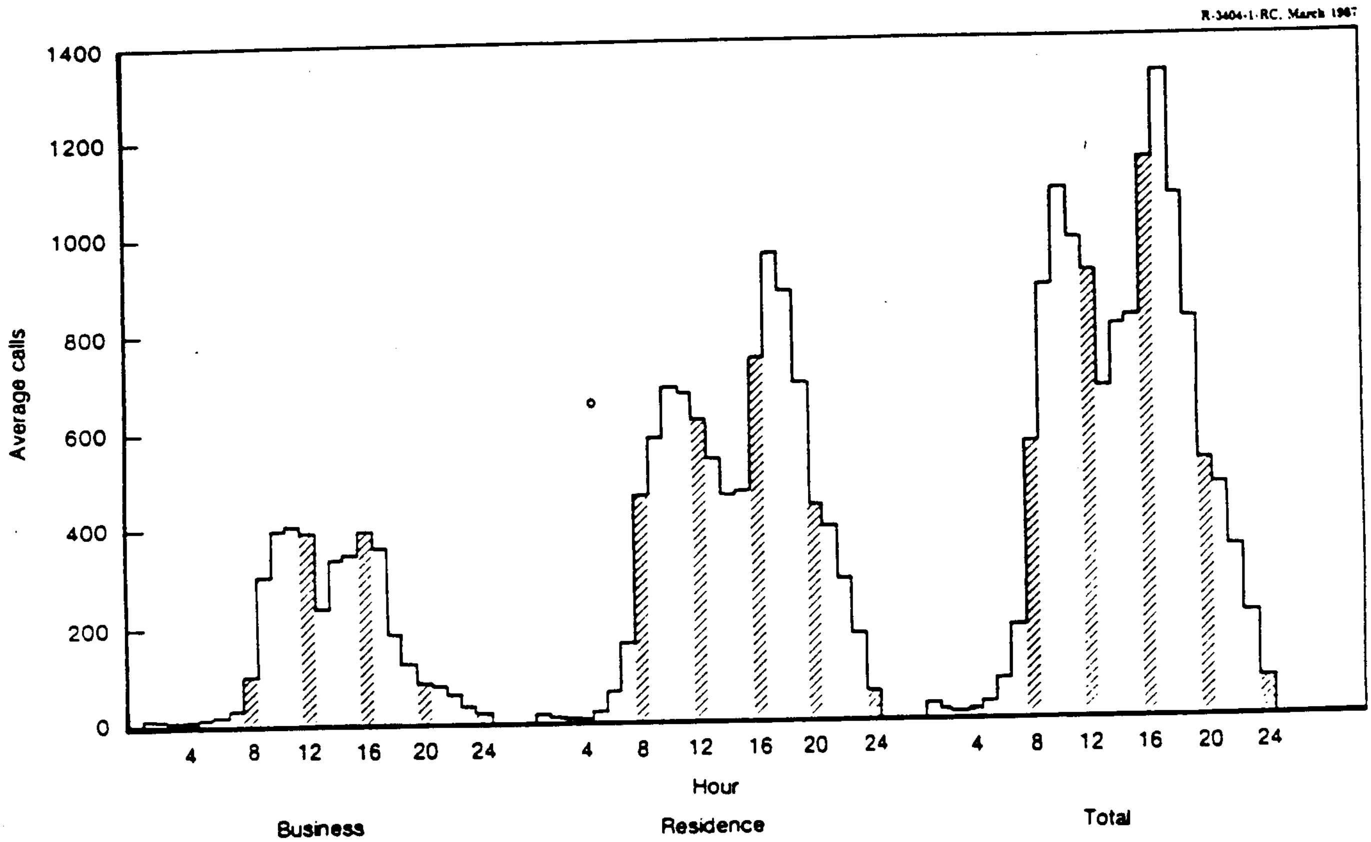


Fig. 8- Average busy-season hourly calls

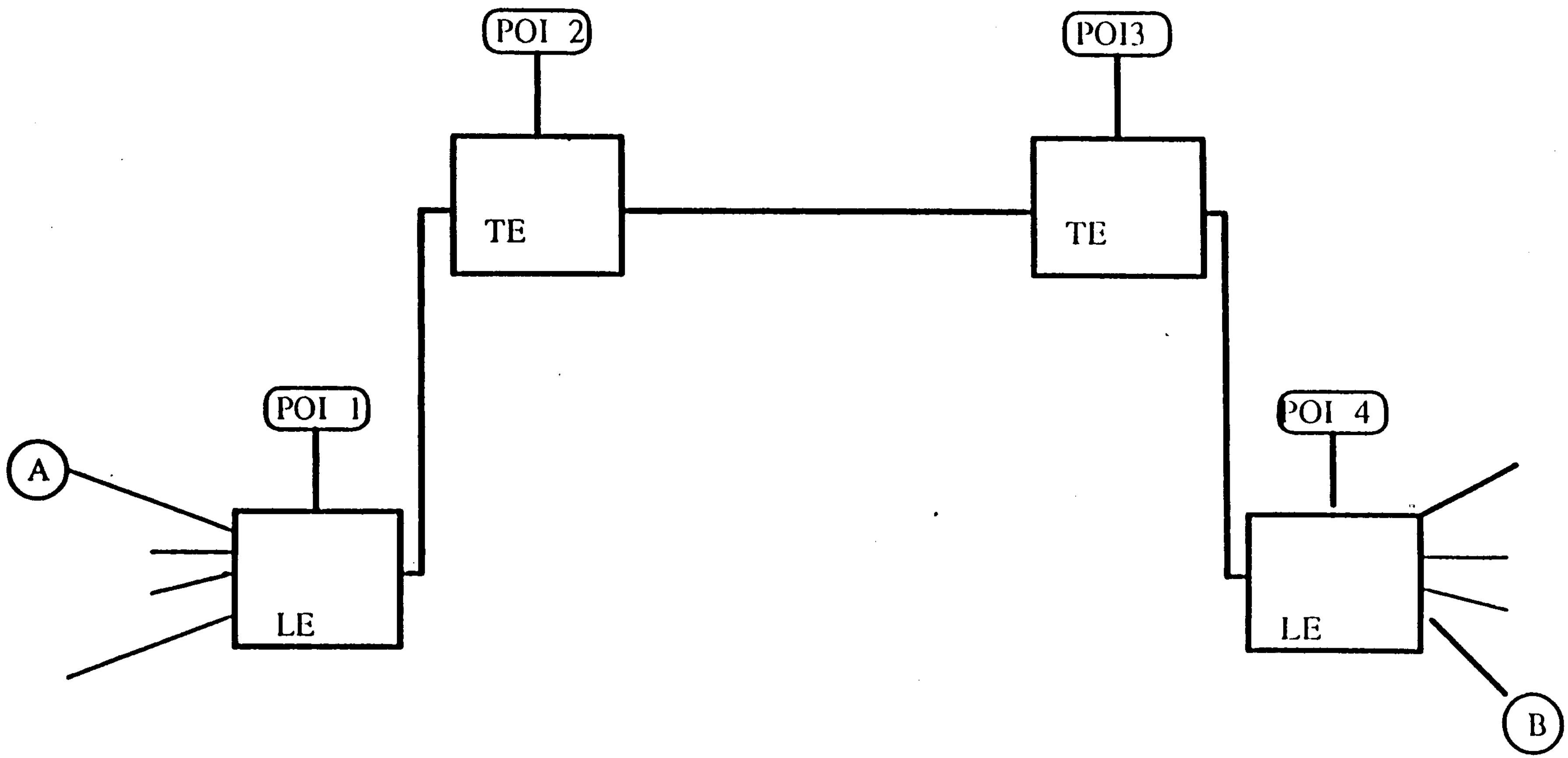


Figure 9