"Models for Planning the Evolution of
Local Telecommunication Networks[†]"

A. Balakrishnan • [††]

T. L. Magnanti • [†††]

A. Shulman ••

R. T. Wong ••• [††††]

MIT Sloan School Working Paper #3025-89-MS

May 1989

*   Sloan School of Management
Massachusetts Institute of Technology
Cambridge, Massachusetts

••   GTE Laboratories
Waltham, Massachusetts

•••   Krannert Graduate School of Management
Purdue University
West Lafayette, Indiana

Abstract

The rapid progress of communications technology has created new opportunities for modeling and optimization of local telecommunication systems. The complexity, diversity, and continuous evolution of these networks pose several modeling challenges. In this paper, we present an overview of the local telephone network environment, and discuss possible modeling approaches. In particular, we (i) discuss the engineering characteristics of the network, and introduce terminology that is commonly used in the communications industry and literature; (ii) describe a general local access network planning problem, and motivate different possible modeling assumptions; and, (iii) summarize various existing and new planning models.

The topics discussed in this paper may be of value to researchers interested in modeling local telecommunications systems, as well as planners interested in using such models. Our goal is to present to modelers relevant aspects of the engineering environment for the local access network, and to discuss the relationship of the engineering issues to the formulation of economic decision models. We indicate how changes in the underlying switching and transmission technology affect the modeling of the local telephone network. We also review various planning models and discuss possible optimization procedures for these problems.

1. Introduction

Over the last three decades, communication network planning and routing has been a fertile problem domain for developing and applying optimization models. Two main driving forces underlie these modeling efforts: (i) the enormous investments in communication facilities (estimated at around $60 billion in 1980 in Bell System transmission facilities alone (AT&T Bell Laboratories, 1986), and over $100 billion in total for the U. S.) offer significant opportunities for cost savings with even modest improvements in the design and operation of communication networks, and (ii) rapid technological and regulatory changes provide novel design alternatives and operating environments. This paper describes several modeling approaches for addressing contemporary design problems that arise in one major component of a telecommunication system, namely, the local access network. As a starting point, the paper first motivates and delineates the context for these problems.

In the next few years, the telecommunications industry – the nature of its services and the volume of its demand – should change radically. Several developments mark the emergence of a new era in communications: replacement of analog transmission by digital technology, decreasing cost and increasing bandwidth of fiber optic transmission equipment relative to conventional copper cables, increasing competition among providers of telecommunication services, and adoption of international Integrated Services Digital Network (ISDN) standards. As ISDN becomes fully operational, and telephone companies complete the transition to digital switching and fiber optic transmission, users will have access to a broad range of new services combining voice, data, graphics, and video. New applications include telemetry, database access, videophone facilities, improved networking services, access to packet networks, and customer controlled network management. Telephone companies are already planning for an even more ambitious expansion of services and capabilities (the so-called broadband ISDN network) when fiber optics will permeate the entire communication system, all the way to the individual customers' homes (Kostas (1984), Dettmer (1985), Toth et al. (1985), *The Economist* (1987), *Fortune* (1988)). Thus, ISDN combined with the new switching and transmission technologies is expected to greatly stimulate network usage, particularly for data and video services, and may even change the rules of competition in industries that depend heavily on information movement.

The investment in new facilities and equipment will be enormous. For instance, in 1986, AT&T and eight other companies reportedly spent about $ 5 billion to build long-distance fiber optic networks for the continental United States (*Forbes*, 1986); moreover, 41 telecommunication organizations in 29 countries planned the world's first intercontinental lightwave (fiber optic) undersea cable system at a cost of $ 1 billion (*Telephony*, 1985). Because switching and transmission equipment are very capital intensive, network planners face complex choices concerning where and when to expand capacity or replace current technology in order to meet the increasing demand for different types of services. The emergence of new technologies, which introduces additional tradeoffs that require new models for analysis, further complicates the planning issues in the industry.

Network modernization and expansion is particularly critical in the local access component of the communication system, both for strategic and economic reasons. In the last few years, the long-distance carriers have almost completed the transition to digital switching technologies, and radio and fiber optic transmission. In contrast, the technological changes in the local telephone network, which accounts for approximately 60% of the total investment in communication facilities, have been much more modest. For instance, in 1987, only 20% of all the local access networks in the U. S. employed digital switching (*The Economist*, 1987). Thus, the ability to offer the proposed advanced ISDN telecommunication services is limited by the current capabilities of local networks, and local telephone companies face competitive pressures to upgrade their networks rapidly. Furthermore, the cost of digital and fiber components (transmission media and electronic switching devices) has been steadily declining relative to analog components (which are also less reliable) and copper cables; this trend is expected to continue, and will make the newer technologies more attractive even on a purely economic basis, particularly for medium and long distance communications.

Given the high costs for complete modernization, and the regulatory uncertainty about the types of services that they can offer, local telephone companies currently have adopted a strategy of investing in expansion and new technology projects primarily based on economic viability (Coathup et al. (1988)). Network planners, therefore, require decision support models that identify cost-effective expansion and modernization strategies. However, the planning models used in

the traditional analog and copper environment are no longer adequate since they do not account for the new alternatives created by technological progress. For instance, deploying concentrators and multiplexers in the local access network now provides an alternative method (instead of cable expansion) for increasing network capacity.

This paper focusses on contemporary evolution planning models for the local access component (from the customer premises to the serving switching center) of public telephone networks. We do not address the design problems for long-distance networks, except to emphasize the differences between local access and long distance network planning. Similarly, our models may not apply directly to data networks, private networks and rural networks since these latter network types employ different technologies (such as radio transmission and packet switching) and different design criteria (e.g., packet delay, routing efficiency, and network vulnerability).

Our purpose in this paper is to discuss alternative **modeling approaches** rather than a specific methodology for local access network planning. The various models that we consider differ in their underlying assumptions, complexity and computational tractability. We focus on **economic models** for aggregate planning (also called *fundamental* planning in the industry) rather than on detailed engineering models of different technologies. Thus, we are concerned with identifying the broad pattern of network evolution, specified by the capacity, location, and timing of investment in different switching and transmission resources. We review some of the underlying telecommunications technology, and contrast the traditional network planning methods developed for the copper and analog environment with the requirements imposed by the newer technologies. We briefly describe solution methods for the different models.

The rest of this paper is organized as follows. Section 2 describes the characteristics of the local telecommunication network, contrasts it with long-distance networks, and introduces some terminology commonly used in the communications industry and literature. This discussion has two purposes: (i) to highlight technological issues that are important in formulating appropriate optimization models, and (ii) to introduce analysts who might not be familiar with the telecommunication industry to some of the prevailing and expected technology. Section 3 presents a formal description of the local access network planning

problem, and motivates the different possible modeling assumptions. This section begins with a general problem description incorporating all the complexities of the design task, and presents a framework that encompasses a wide range of static local access network planning models. We then identify the key dimensions that differentiate various modeling approaches. Section 4 discusses several planning models, and relates each approach to our modeling framework. We first review some models proposed in the literature, and then describe two new models - one using a fixed-charge network design formulation, and another based on tree covering concepts. Section 5 offers concluding remarks.

2. The Local Telecommunication Network

This section describes the typical hierarchical organization of telecommunication networks, and distinguishes the characteristics and design criteria for different levels in this hierarchy. We then describe the local access network in greater detail, trace its evolution over the last few decades, and introduce some communications terminology. This description is not meant to be a complete technical discussion, especially since telephony practices vary from country to country. Our intent is merely to describe important features of each technology so that we can represent them adequately in our economic models.

2.1 The Communication Network Hierarchy

Most national telecommunications networks can be broadly divided into the three main levels shown in Figure 1, namely,

(i) the *long-distance, toll* or *inter-city* network that typically connects city pairs through *gateway nodes* (also called *point-of-presence* nodes, Lavin (1987));

(ii) the *switching center* or *inter-office* network that connects clusters of customers within a city, through nodes called *switching centers* (also called *local exchanges,* or *central offices)*; and,

(iii) the *local access* network that connects individual subscribers belonging to a cluster to the overall system through the cluster's switching center.

These three levels of the communication system hierarchy are distinguished by the processing capabilities and amount of intelligence they contain, the technologies they employ, the services they perform, and their design criteria.

The telecommunication system's basic purpose is to connect geographically dispersed subscribers. To achieve this purpose, the system can conceptually have several possible designs. Some system designs that would be prohibitively expensive include: (i) a system that maintains a dedicated line between every customer pair; (ii) a completely decentralized design, where each node is equipped with the intelligence (i.e, switching and routing equipment) to find its own routes (this network configuration would also pose significant coordination problems, for example, when multiple subscribers request the use of the same transmission facilities); and (iii) a very centralized system connecting every user to a centralized switching facility (by a dedicated line) that makes required connections between incoming and outgoing lines.

The three level system that has evolved over the years is a combination of these extreme designs. At the lowest level of the hierarchy, the end nodes (corresponding to the individual subscribers) do not contain the intelligence to perform any switching functions. Instead, the *local access network* contains a dedicated communication channel connecting each customer to a switching center which serves as the interface to the higher level telecommunication network. Typically, this local access network has a tree configuration. Currently, most (approximately 80% in the U.S.) local access networks use analog transmission on copper cables, and do not contain electronic devices, called multiplexers or concentrators, that enable users to share transmission lines by compressing traffic.

The *inter-office network* interconnects all the switching centers within a restricted geographical region (for example, within each city) via high speed transmission lines and possibly through tandem switches; it also provides access to the nearest gateway node of the long-distance network. The number of switching centers in each region depends upon its area and population density. The inter-office network contains limited intelligence for routing incoming messages to the

appropriate downstream switching center or gateway node. Increasingly, this subnetwork employs digital switching and transmission technologies using high speed copper and fiber optic media, and provides alternate paths for communication between switching centers.

The gateway nodes serve both as interfaces between local switching centers and the *long-distance network*, and as tandem or intermediate switches for transshipping messages between other city pairs in the long-distance network. (Some cities might contain more than one gateway node.) Typically, the gateway nodes contain intelligent hardware that performs switching, traffic compression (concentration), and some service functions (such as directory assistance). The long-distance network has a relatively dense topology, and therefore provides multiple communication paths between each origin-destination pair. Currently, the long-distance networks in the U.S. are almost completely digitized, and employ high frequency transmission using fiber optics, microwave (radio), and satellite communications.

To better understand the role of the three levels in the communication hierarchy, let us consider the routes used when two customers communicate with each other. If both customers are connected to the same switching center, then the transmission path consists of the channel from each customer to the switching center, where they are connected (either mechanically or electronically) to each other. Thus, the customer-to-customer transmission does not require any communication resources outside the local access network. Consider now the case when the two customers are connected to different switching centers located in the same geographical region. The communication path then consists of the channel from each customer node to its switching center, and a route on the inter-office network connecting the two switching centers (possibly through some intermediate switches). Finally, for two customers located in different geographical regions, the communication path uses all three levels in the hierarchy - the respective local access networks, the inter-office network connecting the respective switching centers to the gateway nodes, and the long-distance network between the gateway nodes.

Ideally, the design of a telecommunication network should simultaneously account for all three levels of the network hierarchy since the capacity requirements at the different levels are interdependent. For instance, the number of customers

assigned to each switching center, and their respective communication requirements (i.e., to whom they communicate and how often) would determine the desired switching capacity at the switching center as well as the transmission capacity for the inter-office network. Similarly, the capacity of the long-distance network depends on the assignment of switching centers to each gateway node. However, in practice, analysts typically plan each level separately. This decomposition approach is appropriate for several reasons:

(1) Different corporations or agencies might own and operate the different levels of the network. For instance, in the U. S., different regional operating companies (or local exchange carriers) own and operate the local access and inter-office networks, while several independent carriers compete in the long-distance business.

(2) Even when one company owns all three components of the network, solving a single monolithic model that simultaneously accounts for all the levels is almost impossible because of its very large dimensions (measured in terms of the number of nodes, links, and origin-destination demand pairs).

(3) Different levels of the network employ different technologies (e.g., digital versus analog switching, fiber optic and satellite transmission versus copper cables), have different transmission speeds, and, more importantly, have different design criteria, planning philosophies, and priorities.

For instance, in the local access network, each customer has a dedicated channel to the switching center. On the other hand, users share the switching and channel capacities in the inter-office and long-distance networks, i.e., the system dynamically assigns channel capacity as the need arises for communication between various origin-destination pairs. Capacity utilization, blocking probabilities, reliability (which might be addressed by providing alternate transmission paths) are important criteria for the design of inter-office and long-distance networks; these criteria are currently not as critical for the local access network. As new technologies and user services emerge, reliability issues may become important in the local access networks as well.

In this paper we focus on decision models for planning the evolution of local access networks. To understand this planning task, we first describe the local access network in greater detail.

2.2 Evolution of the Local Access Network

The local access network (also called the *outside plant, local loop, subscriber loop* or *local exchange network*) connects individual customers to a switching center. Like the overall communication system, this network also has a hierarchical structure; the industry often refers to the three levels as *routes, feeder networks,* and *distribution networks.*

A *route* is a portion of the local access network; it is a subnetwork interconnecting all customer nodes that communicate with the switching center via a particular link incident to it. Figure 2 shows a typical route of a local access network; each switching office may serve as the termination point for 3 to 5 routes (Koontz (1980)). The Bell System contains around 40,000 such routes (Ciesielka and Douglas (1980)).

Each route is in turn divided into two segments: the *feeder network* connecting the switching center to intermediate nodes called *distribution points* (or *control points*), and *distribution networks* connecting each distribution point to the customer premises. The *feeder network* consists of cable groups that are either buried, installed in ducts, or mounted on poles, and are accessible at intermediate points. The segment of cables between two adjacent distribution points along the route is often called a *feeder section*. The feeder network has a tapering structure, i.e., the number of cables in each feeder section decreases as we move away from the switching center. The *distribution network* taps into the feeder network via lateral cables at the distribution points. The area served by a distribution network (i.e., the area assigned to each distribution point), sometimes called an *allocation area* (Gibson and Luber (1980)), typically has a diameter of a few thousand feet. The number of distribution points served by a switching center varies from 20 to 200 and each distribution point might serve as many as 500 customers. Most feeder and distribution networks have a tree structure that provides a unique transmission path from each customer to the switching center. (See Griffiths (1986) for a more comprehensive description of local telecommunication networks.)

Traditionally, for capacity planning, the distribution network is designed for ultimate demand (which is relatively small) in order to avoid subsequent disruption of service for laying new cables and to exploit economies of scale. On the

other hand, a feeder network that is designed for ultimate demand will incur significant investment and very low cable utilization rates. Thus, planners design the feeder network's capacity to meet only the medium-term (typically, 3 to 7 year) demand, and periodically reinforce the network by rearranging and adding cables (Ciesielka and Long (1980), Elken (1980), Friedenfelds and McLauglin (1979)). In this paper, we consider the medium-term feeder capacity planning problem. We next trace the evolution of technologies and planning practices in the feeder network. From a modeling perspective, we might classify the technological developments into three stages.

<u>Stage 1:</u> *The basic feeder network*

The basic feeder network employs analog transmission at the voice frequency of 4 Khz over copper cables (twisted wire pairs). Each feeder section contains cables of varying gauges (companies in the U.S. use four different wire gauges), with the gauges becoming progressively coarser as the distance from the central office increases. Each customer is connected to the switching center using a dedicated line. Physically, the line for a customer might consist of wire segments (possibly with different gauges) belonging to each downstream (i.e., towards the switching center) feeder section, that are joined at the intermediate distribution points. Initially, the industry used electromechanical switches (step-by-step and cross-bar exchanges) to make the line connections at the switching center. Increasingly, it is now replacing these switches by more reliable and economical digital devices; with the progressive digitization of inter-office networks, these digital switches also serve as analog-to-digital interfaces (between the analog local network and the digital inter-office network).

In this setting, one of the main design concerns is to provide acceptable transmission quality by ensuring that the circuit connecting each customer to the switching center satisfies the maximum permissible wire resistance (around 1300 ohms, increasing to 2500 ohms with *range extenders*). Thus, the network engineering task consists of determining the combination of wire gauges to use in each feeder section so that every customer is served by a circuit that satisfies the maximum resistance requirements. This procedure is sometimes called Resistance Design (Ciesielka and Douglas (1980)). Often, the feeder sections contain unallocated or spare cables that may be subsquently utilized to accommodate demand growth and subscriber movement.

Observe that the basic feeder network can respond to increased telecommunication demand only by adding and reassigning cables within each feeder section. Planners sometimes refer to this method as *physical pair facility relief*. If the feeder network has a tree structure, the demand (i.e., number of lines required) at the various distribution points uniquely determines the cable capacity requirements in each feeder section. Thus, if the total demand increases or if customers move from one allocation area to another, the cable requirement for each feeder section changes. Any section with cable demand in excess of availability is said to have *exhaust*. The feeder planning exercise then considers two strategies: (i) feeder cable reallocation, and (ii) feeder cable expansion.

For a given level of projected short or medium term demand at each distribution point, *feeder reallocation* methods attempt to identify a feasible reassignment of currently allocated and spare feeder cables within each section to various downstream distribution points in order to delay cable expansion. Gibson and Luber (1980) describe a heuristic allocation method that sequentially considers different feeder sections in decreasing order of criticality (which is defined in terms of the earliest time at which the section experiences exhaust). For each section, the method allocates spare capacity to minimize the total cost of rearrangement and blocked service requests. This procedure applies to both high growth and low growth/high movement routes. Elken (1980) formulates the feeder reallocation task as a separable convex programming problem, and proposes an iterative procedure which solves a sequence of linear programs. The model considers all sections simultaneously, incorporates the different cable gauge requirements, and minimizes the operating costs due to subscriber movement plus the relief and rearrangement costs. The method extends to a multi-period setting where the user can constrain the cable allocations in order to avoid frequent rearrangements within the planning horizon.

Feeder expansion models (e.g., Freidenfields and McLaughlin (1979) and Koontz (1980)) attempt to determine the number of additional cables to install in every time period (typically, every year) of the planning horizon in order to relieve the projected exhaust at minimum total *present worth of actual costs*. As mentioned previously, when cable expansion is the only available exhaust relief method, and if the local access network has a tree structure, each feeder section can

be analyzed independently for capacity planning purposes. Thus, physical pair facility relief models used in this context do not incorporate any spatial coupling between sections. Friedenfelds and McLaughlin (1979) present a multiperiod capacity expansion model that considers each section separately, and incorporates detailed engineering specifications such as different cable gauge requirements, discrete cable sizes (i.e., the number of twisted pairs in each cable) and cable space availability in the existing infrastructure. The model also permits replacement of existing cables with larger ones to avoid building new conduits. The authors propose a heuristic method based on an enumeration tree that approximates the discounted cost of completing the decision sequence beyond the first two installation/replacement decisions.

Stage 2: *Feeder networks with remote electronics*

From a modeling point of view, the next major stage in local network evolution occurred when the communication industry developed *pair gain* devices, i.e., *multiplexers* and *concentrators*, for use in the local network. A multiplexer is an electronic device that compresses or interleaves signals from several incoming lines into a composite outgoing signal that has a higher frequency but requires only a single line (or a pair of lines). The system assigns each incoming signal to a separate 'channel' in the combined outgoing transmission. (Channels correspond to preassigned non-overlapping frequency bands for frequency division multiplexing, while they correspond to time slots in time division multiplexing.) We refer to the ratio of input to output signal frequencies as the *traffic compression* ratio (also called the *multiplexing* ratio). Like multiplexers, *concentrators* also perform traffic compression, transforming multiple incoming signals into a single outgoing high frequency signal. However, the output signal from a concentrator does not have a dedicated channel for each input line. Concentrators have fewer output channels than the total number of input lines; these channels are dynamically assigned to the input lines as the need arises. The ratio of incoming lines to outgoing channels is called the *concentration ratio*. Planners employ traffic engineering methods to select proper concentration ratios based on desired service levels (specified in terms of maximum permissible blocking probabilities).

In local access network applications, multiplexers and concentrators enable multiple users to share the same physical line on the feeder network. In particular, these devices are located at distribution points where they combine signals from

several distribution cables (from the customer) onto fewer outgoing feeder lines (to the switching center). Thus, multiplexers and concentrators provide an alternative to cable expansion for relieving exhaust when demand increases. These remote electronic devices are available in several configurations, with varying input capacities (i.e., number of input lines) and different traffic compression ratios ranging from 2:1 to as high as 96:1.

While multiplexing/concentration reduces the number of cables required in downstream feeder sections, these cables must now handle higher frequencies, since the transmission frequency of the output signal increases in direct proportion to the traffic compression ratio. The telecommunications community has adopted a set of standard frequencies for local network transmission. Conventional copper cables (twisted pairs) have a limited bandwidth (around 150 Khz); at higher frequencies, the quality of the signals deteriorate rapidly because of attenuation. To handle higher frequency signals, either the copper cables must be enhanced (or *conditioned* or *groomed*), for instance, by adding intermediate repeaters, or they must be replaced by coaxial or fiber optic cables. Often, these enhanced transmission media can be installed in existing ducts containing copper cables.

The introduction of digital transmission *within the local network*, and switching capabilities at distribution points further changes the design requirements. Increasingly, the communications industry is digitizing local access networks for compatibility with ISDN standards; furthermore, digital hardware is more reliable, and easier to maintain and upgrade (Combot and Epstein (1979)). With digital transmission in the local loop, the electronic device at the distribution point serves both to compress traffic and to convert analog signals to digital signals. Correspondingly, the analog-to-digital conversion at the switching center interface with the interoffice network is no longer necessary. The switching function is also becoming increasingly decentralized with the use of *remote switches* (located at distribution points) that also perform concentration and analog-to-digital conversion. Two customers connected to the same remote switch can directly communicate through this switch instead of using the channel capacity to and from the switching center. Thus, the remote switching facility further reduces the downstream transmission requirements, and serves as an additional (though limited) means to relieve exhaust in the feeder network.

As the previous discussion suggests, multiplexers, concentrators and other remote electronic devices have created alternative methods for responding to increasing telecommunication demand. With these additional options, the designer must consider various choices for locating, sizing, and timing the installation of remote electronics (multiplexers, concentrators, switches), in addition to the conventional options of increasing cable capacities in different feeder sections. Furthermore, unlike the older technologies, these new devices introduce spatial couplings, i.e., we can no longer consider each feeder section in isolation since increasing demand at upstream distribution points does not necessarily translate into increased cable capacity requirements on every intermediate feeder section since we can deploy remote concentrating devices.

Stage 3: *Fiber in the local access network*

Currently, much of the traffic over public local access networks is voice communication (and a limited amount of data transmission), and except for a few experimental networks, telephone companies have not deployed fiber optic (or lightwave) transmission in the local loop. When the industry introduces lightwave transmission on a regular basis, additional design issues will arise. Lightwave transmission facilities consist of a pair of *fiber optic terminals* (or *fiber terminating equipment*) connected by a fiber cable. The fiber optic terminals convert electrical (analog or digital) signals into very high frequency optical signals, and may perform optical coupling and multiplexing functions as well. Fiber cables have a very large bandwidth (some with a transmission capability of over 1 Tbs, which is effectively unlimited for local network applications) and can, therefore, accomodate a large number of multiplexed channels; indeed, the electronics in the fiber terminating equipment is currently the main limiting factor for the number of channels that can be multiplexed on fiber. For local network applications, the cost of fiber terminating equipment is expected to dominate the fiber cable costs (especially, if fiber cables are installed in existing underground ducts) due to the relatively short distances between the distribution points and the switching center.

The use of fiber optics in the local network is still in the developmental stage. Hence, the characteristics and capabilities of the technology, and even the network and device configuration plans are constantly changing (Anderson (1988), Carse (1986), Ensdorf et al. (1988), Toth et al. (1985), Snelling and Kaplan (1984)). Several competing network topologies have been proposed in the literature for fiber-based

local networks (see, for instance, Campbell (1988), Garbanati and Palladino (1988), Sirbu and Reed (1988), White (1988)). For instance, the double star configuration contains fiber optic terminals at distribution points that are connected to terminals at the switching center via dedicated 'umbilical' fibers. Alternatively, if fiber is also installed in the distribution network, the *customer premises equipment* might itself contain fiber optic electronics, obviating the need for electrical-to-optical conversion at the intermediate distribution points or switching centers. In this latter scenario, the distribution points might possibly contain optical concentrators that combine several incoming optical signals into a single outgoing optical signal that is transmitted on a single fiber cable to the switching center. If the cost of fiber cables declines, then installing a dedicated fiber from the customer premises to the switching center (without intermediate optical concentrators) might become a cost-effective strategy as well.

The development of fiber splicing devices exemplifies the rapidly changing technology and its planning implications. Until recently designers could not splice or join fiber cables at intermediate distribution points. Thus, planners essentially considered the fiber optic cable and the terminals connecting each distribution point to the switching center as an integral unit, and did not plan fiber capacity section by section (as they do for copper cables). Some ongoing research and development efforts have demonstrated the technical and economic viability of fiber splicing devices. This development creates new opportunities to exploit economies of scale by sharing fiber cables among several distribution points.

In this paper, we will argue that, for planning purposes, fiber optic terminals essentially act like concentrators with very high traffic compression ratios. Thus, we do not represent the unique characteristics of fiber optic transmission in great detail, particularly since this technology is still evolving. When the technology develops further and telephone companies gain experience with its deployment, network planners might require more sophisticated models that distinguish fiber optic transmission from conventional electrical transmission.

In the next section, we formalize the feeder network planning problem that we wish to address, and motivate several modeling assumptions; we will subsequently use these assumptions to differentiate various possible modeling approaches.

3. Local Network Planning: Problem Definition and Modeling Assumptions

This section presents a general framework and a formal definition of the local access network planning problem, defines the scope of models that we will subsequently consider, and identifies some key assumptions that distinguish different modeling approaches for this problem. Before restricting the focus for the paper, we will first describe the planning task in its most general form.

The feeder capacity planning exercise begins with a forecast of telecommunication demand at each distribution point for the duration of the planning horizon. The demand projection is based on information about new construction and customer movements in the allocation areas served by the network, followed by a traffic engineering phase that translates individual customer requirements into equivalent demands at each distribution point. The basic unit of demand for voice transmission is a *circuit*. For analog transmission, each circuit represents a bandwidth requirement of 4 Khz, and requires one twisted pair of copper wires. The corresponding digital equivalent in the U.S. is the DS0 signal which has a transmission rate of 64 Kbps (Kilobits per second). The demand for data, video, and other wideband services is usually expressed as a multiple or fraction of the basic DS0 rate (for example, the DS1, DS2, and DS3 rates are, respectively, 24, 96, and 672 times the DS0 transmission rate, while data transmission rates may be 9.6 Kbps or lower).

Consider a medium-term planning horizon during which the demand for different services - voice, data, video, telemetry - is expected to grow. In general, different services may have different transmission and processing requirements within the local access network. For instance, services such as pay-per-view video facilities and directory inquiries must be processed differently than regular voice communication; similarly, two-way videophone services require extremely high bandwidths. In addition to incorporating the different processing requirements, the general planning model must also account for the different transmission media (twisted wire pairs, repeatered copper cables, coaxial cables, and fiber optic cables) that each service type and traffic processing device requires. For example, video signals cannot be transmitted over twisted pair copper cables; similarly, a concentrator that transmits output signals at the DS2 rate requires enhanced copper, coaxial, or fiber optic media. Furthermore, the model must incorporate various

technological and policy restrictions on the design, such as providing multiple paths to some preferred large-volume business customers, or limiting the distance between each customer and the nearest concentrator or remote switch (which we refer to as a *proximity restriction*). For instance, data transmission technology limits the maximum copper cable length from the customer to the first electronic device to around 12 kilofeet; beyond this distance, data transmission exhibits rapid degradation. Proximity rules may also stem from strategic considerations. For example, telephone companies may adopt a policy of positioning all new fiber terminating equipment close to the customer in anticipation of future services that require an extension of the fiber network to customers' homes (the so-called *last mile* fiber installations); some companies currently use a planning rule specifying that each customer must be served by a fiber terminal that is no more than 2 to 5 kilofeet away. Finally, the objective of the planning exercise might be either to satisfy all the projected demand at minimum total discounted cost, or to selectively satisfy demand to maximize the total profit. Thus, the general local access network planning problem may be stated as follows:

The given data are:
(1) the demand for each service at every distribution point in each year of the planning horizon,
(2) the processing requirements for different service types,
(3) the revenue per unit of service (only required for profit maximization objective),
(4) the current network configuration, and current processing and transmission capacities (e.g., the number of cables in each feeder section, and the location and size of existing concentrators, etc.),
(5) the installation and operating costs for different possible network enhancements (i.e., addition/expansion of transmission media and nodal processing facilities),
(6) technological and policy restrictions on the local access network design, and
(7) the discounting factor (that may vary by equipment and investment type) to use for computing discounted costs and revenues.
The optimal evolution plan should specify
(a) the location, timing and sizing of various network enhancements, and
(b) the routing of traffic from each distribution point to the switching center.
The optimization objective might be either to

(a) minimize the total present worth of actual costs to satisfy the demand for
each service type in each year of the planning horizon, or

(b) maximize the net present value of profit (= revenue - cost); in this case, the
plan should also specify the proportion of demand for each service type to
satisfy at every distribution point during each year of the planning horizon.

The evolution plan must also implicitly specify the overlay and replacement
strategies for new technologies. Consider, for instance, the introduction of digital
switching and transmission in a local access network that currently employs only
analog transmission. (High traffic compression rates are possible only with digital
electronics; thus, selecting a concentrator with high traffic compression ratio implies
the use of digital technology.) In a multi-period framework for modernizing this
network, the modeler must decide whether the new digital technology will
completely replace the analog technology as soon as it is installed, or whether the
digital technology will initially overlay the analog technology. The latter strategy
staggers the replacement of analog transmission, i.e., the digital technology initially
accomodates only the growth component of demand, and subsequently replaces the
current analog transmission as well. Combot and Epstein (1979), Combot and
Mason (1979), Combot et al. (1981), and Hoang and Lau (1984) have considered such
overlay and replacement tradeoffs for inter-office network planning.

As this problem statement suggests, the general local access network planning
model is extremely complex, has a very broad scope, and is highly intractable from a
model solution perspective. Depending on the specific application context, various
simplifying assumptions might apply. We restrict the scope of our discussions by
considering the following simpler model.

First, we consider only the cost minimization form of the planning problem,
where the entire projected demand at each distribution point must be completely
satisfied, and the objective of the network planning exercise is to minimize the
present worth of all investment and operating costs. Second, we focus on *static*
rather than *multi-period* models. Ideally, demand is specified during each year of a
multi-year planning horizon, and the network evolution plan should accommodate
the demand growth during each year. This multi-period model would account for
the temporal couplings caused by economies of scale, i.e., the optimal investment
strategy might install excess capacity during one year in anticipation of higher

demand in subsequent years. However, multi-period models are much harder to solve compared to a static (or single period) model that only seeks to satisfy demand in the terminal year of the planning horizon. Furthermore, studying static models might possibly give us insights about the more general problem, and the single-period solution algorithms might serve as building blocks for multi-period versions (see, for example, Shulman and Vachani (1988)). Or, as Minoux (1987) has proposed, the static model might be used to first identify the final target network; a subsequent multi-period model would then determine the evolution of the existing network toward the target. Our third simplification concerns the different service types (voice, data, video). We assume that, within the local access network, the various services do not require different processing steps (e.g., a special database query at a nodal processor). Effectively, we ignore any unique information processing requirements, and consider only common traffic processing options. Finally, we do not consider certain policy restrictions, such as providing multiple communication paths to selected customers, since these are company-specific; however, some of our models can account for proximity restrictions.

All the modeling approaches that we discuss in Section 4 make these assumptions, i.e., they are static, cost minimizing, models incorporating only traffic processing options that are common to all service types. As the following detailed problem description suggests, even with this restricted scope, the network planning problem is very complex. The various models that we discuss in Section 4 differ in terms of several additional assumptions that we describe later. Section 5 of the paper discusses model extensions to multiple periods, and mutiple service types. Next, we develop a modeling framework, define the network planning problem more precisely, and introduce some notation. Our framework employs some fairly general modeling constructs, and encompasses a wide range of existing and proposed transmission and processing technologies, cost structures, and topologies.

3.1 Modeling Framework

The local access network planning problem is defined over an undirected network G:(N,A) whose nodes N = {0,1,2,...,n} correspond to the switching center (node 0) and the distribution points (nodes 1 to n) assigned to it. The set of edges E contains edge (i,j) if the local access network currently contains or can potentially contain a feeder section between nodes i and j.

We use the index s for the different services provided by the local access network. For each service type s, let r_s denote the transmission frequency or rate at which the service originates; r_s is also the minimum required transmission rate for service type s. For instance, video signals may require a minimum transmission rate of 1.5 Mbps (the DS1 rate); these signals may be multiplexed to higher rates, if necessary. We refer to the rate r_s as the *basic rate* for service type s. Let d_{is} represent the terminal year demand for service of type s at node i. This demand is expressed in terms of the number of required channels at the basic rate r_s, i.e., the final design should provide d_{is} channels (at rate r_s), for each service s, from node i to the switching center.

As the discussion of Section 2 illustrates, the network planner has two basic strategies for accommodating this projected demand: increasing the transmission capacity in one or more feeder sections, and/or installing traffic compression devices, such as multiplexers, concentrators, remote switches, or fiber optic devices, at the distribution points. Henceforth, we will refer to any device that compresses traffic as a *processor* (or *traffic processor*, or *nodal processor*); we assume that these devices can only be located at prespecified nodes of the network.

Before describing other model parameters, we must first establish the relationship between transmission rates, traffic processors, and transmission media. The local network employs a finite set of (digital or analog) transmission rates or frequencies. Let L = {1,2,...,|L|} denote the index set of available rates, with higher indices *l* corresponding to higher rates r_l. This set includes all the basic rates corresponding to the various service types; in general, it may also contain other frequencies that do not represent basic service rates. Without loss of generality, we assume that each transmission rate has an associated service type; thus, we will use the index s and *l* interchangeably when we refer to service types.

For each transmission rate, we associate a 'preferred' transmission medium (such as twisted wire pairs, repeatered copper cables, coaxial cables, and fiber optic cables). In theory, the designer may have a choice of several alternative media to carry a particular transmission rate (especially if the rate is at the low end of the frequency spectrum); we assume that one of these media is preselected for each rate. Conversely, each transmission medium has a preferred (frequency) range of operation. The ranges for the different media are nonoverlapping and exhaustive (i.e., together, they cover all the possible transmission rates). Thus, the same physical medium may accommodate several adjacent transmission rates.

As we explained previously, traffic processors combine several incoming (lower frequency) signals into a single outgoing (higher rate) signal. The economic models that we consider will differentiate the various processor types in terms of their respective input rates (i.e., frequency of input signals), output rates, and conversion ratios. We do not model other detailed technological differences. By *conversion* ratio we mean the ratio of number of incoming *lines* (e.g., copper wire pairs) to outgoing *lines* (including spare lines for contingencies). (The industry uses a related measure called the *pair gain ratio* defined as (number of input lines - number of output lines) ÷ number of output lines.) This conversion ratio may differ from the ratio of output *rate* to input *rate* (which we call the *traffic compression* ratio) because of differences between the number of input and output channels (the number of output channels is smaller for concentrators), and provisions for spare outgoing lines.

To illustrate these concepts, let us consider a specific commercially available traffic processing device, namely the SLC-96 system (Ciesielka and Douglas (1980)). The SLC-96 is a modular digital carrier/concentrator system that was introduced in the Bell System around 1979. Each module supports 96 voice frequency input lines. The input signals are analog; the SLC-96 system converts them into digital signals before retransmission. For purposes of illustration, we will treat the input rate as the (digital) DS0 rate (64 Kbps). The system performs two-to-one digital concentration, i.e., the number of output channels is half the number of input lines; thus, each module has 48 output channels. These 48 output channels are transmitted over two standard T1 digital lines, each carrying 24 channels. The system also requires a spare T1 line to assure continuity of service when one of the

main T1 lines fails. Each T1 line might consist of two pairs of copper wires, with intermediate repeaters; the transmission rate over each line is 1.536 Mbps (= 24 channels x DS0 input rate of 64 Kbps), which corresponds exactly to the DS1 rate. Thus, the SLC-96 has a concentration ratio of (96 input channels + 48 output channels) = 2, a traffic compression ratio of (1.536 Mbps + 64 Kbps) = 24, and a conversion ratio of (96 input pairs + {3T1 lines x 2pairs/line}) = 16. One version of the SLC-96 system permits stacking of up to ten modules (providing service for up to 960 customers).

The SLC-96 has a DS0 input rate, a DS1 output rate, and a traffic conversion ratio of 16. Other processing devices might have different input rates, output rates, and/or traffic conversion ratios. We define each available combination of these three characteristics as a distinct *processor type*; thus, the same family of devices (such as digital multiplexers) might contain several different processor types. Let M = {1,2,..., | M |} be the index set of the available processor types. For processor type m, let $l_1(m)$ and $l_2(m)$ (> $l_1(m)$) denote, respectively, the indices of the input and output rates, and let ρ_m represent its traffic conversion ratio. Effectively, the input and output rates determine the type of input and output media that the processor requires, while the conversion ratio determines the number of physical lines of the output medium required per incoming line.

Processor capacities are specified in terms of their maximum number of input lines. We denote the current capacity of a type m processor located at node i as A_{im}. Installing new processors or expanding existing capacities entails *processor costs* which may vary by location, processor type, and the required additional capacity. Let $h_{im}(y)$ represent the cost of installing (expanding) a processor of type m with capacity y at node i. For instance, this cost function might consist of fixed costs (for acquiring land, constructing buildings, pedestals, cabinets and other infrastructure), and variable or volume-dependent costs (e.g., for each module) that depend on the desired capacity. We expect these processor cost functions to be concave (as a function of capacity) because of economies of scale. Next, we discuss the representation for transmission facilities, and propose a general classification of transmission media to model various possible transmission cost structures and topologies.

We will differentiate transmission facilities according to the *medium* or *cable type*. Let K = {1,2,..., |K|} be the index set of available transmission media. Each medium can accommodate a (nonoverlapping) range of transmission rates. Let L(k) denote the index set of transmission rates that use medium k. To describe the transmission cost structure (i.e., the cost of installing or expanding cable capacities), we classify cable types into two categories: *sectional* cables and *continuous* cables. Sectional cables are planned and installed in sections. For instance, conventional copper cables belong to this category; they can be spliced and joined at the intermediate distribution points. The cost associated with expanding sectional cable types is the sum of the cable expansion costs in each feeder section. In contrast, for certain other cable types, technological and economic considerations might require that we use a continuous or 'umbilical' connection from the distribution point to the switching center. We refer to these cable types as *continuous* cables; typically, they serve as the output media for traffic processing devices. Until the recent development of fiber splicing devices, fiber cables belonged to this category. Observe that sectional and continuous cables share the same infrastructure (underground ducts, poles, etc.). Also, some cable types might possibly be used in either sectional or continuous mode depending on the transmission rate and the traffic processing devices; in this case, we consider the sectional and continuous versions of the same medium as two different cable types. Let K' and K" denote, respectively, the index sets of sectional and continuous media. For each sectional cable type k, let B_{ijk} denote the current capacity (in number of lines) of that cable type on feeder section (i,j). Similarly, for each continuous cable type k, let B_{ik} be the number of type k cables connecting node i to the switching center. We next discuss the possible cost structures for expanding these transmission capacities.

In general, the total (transmission) cost to expand cable capacities may not be completely separable by cable type; rather, this cost consists of two components: a separable cost component pertaining to each *individual* cable type, and a *joint* or shared cost that is common to several (or all) cable types. Joint costs arise because different cable types share the same infrastructural facilities. For instance, once we incur the costs for building underground ducts, several different media can share these ducts. Let $f_{ijk}(x)$ denote the (individual) cost of expanding the capacity of sectional cable type k on section (i,j) by x units; similarly, let $f_{ik}(x)$ be the cost of adding x units of continuous cable type k from node i to the switching center. Because of economies of scale, both these cost functions are likely to be concave, and

- 22 -

might consist of, say, fixed as well as variable costs that depend on the volume of traffic. Finally, let $g_{ij}(x_1,x_2,...,x_{|K|})$ represent the joint cost of installing x_k type k cables, for $k = 1,2,...,|K|$, along section (i,j).

The (static) local access network planning problem involves minimizing the sum of the individual and joint transmission costs, and the total processor costs in order to meet the projected demand. The optimal plan must specify the following decisions:

(a) where to locate processors;

(b) for each selected location, what type of processor to use, and with what capacity;

(c) where to expand or install new transmission (cable) capacities, and by how much; and,

(d) how to route the demand from each distribution point to the switching center, i.e., where to process the traffic from each node, and which feeder sections to use.

In order to determine the best plan, the local access network planning model must address two basic tradeoffs: a tradeoff between increasing transmission capacity on one or more downstream sections (i.e., feeder sections en route to the switching center) versus installing nodal processors, and a tradeoff between using several decentralized processors instead of a few high capacity processors. Before presenting a mathematical programming formulation for the local network planning model, we first illustrate these tradeoffs using a simple example.

3.2 Example

Figure 3a illustrates a local access network problem with a single service type (with a basic rate of r_1) and single medium. The given network has a tree structure, consisting of a single (sectional) medium (say, copper cables) with existing capacities as shown in the figure. However, this capacity is inadequate for the projected demand level. The feeder sections with projected exhaust (i.e., capacity shortfall), which we call *bottleneck* edges, are highlighted with heavy shaded lines. The projected exhaust represents the amount by which cable capacities must be expanded in a conventional physical pair relief strategy that does not consider traffic

processing options. Figure 3b shows the cost function for expanding cable capacity along feeder section (i,j); it consists of a fixed charge F_{ij} plus a (constant) per unit cost c_{ij}. In this example, the current network does not have any processors. To relieve exhaust, the planner has available a single processor type (say, type 1), with a traffic conversion ratio ρ_1 of 10. This processor accepts input signals at rate $l_1(1) = r_1$ (i.e., input rate = basic rate of the single service type), and has an output rate of $l_2(1)$. Observe that by considering only a single processor type, we automatically preclude multiple processing steps in series. Assume that the processor is available in three sizes (capacities), each with an associated fixed cost. Thus, the processor cost, as a function of its capacity, is a step function as shown in Figure 3c. We consider two possible transmission rates: the basic rate r_1 of the single service type (without any traffic compression), and a compressed rate $l_2(1)$ corresponding to the output signal of the traffic processor. We assume that the compressed rate requires the same medium (say, copper cables, possibly *groomed* or conditioned to handle the higher rate) as the basic rate signal.

Figure 4 shows one possible expansion plan for the network example of Figure 3. This plan entails installing a processor with a capacity of 400 units at node 5, and expanding the cable segment along section (3,1) by 100 units. The processor at node 5 processes all the traffic from nodes 2, 4, 5, 8, and 9. Its output signal, shown in dotted lines, travels from node 5 to node 0 (the switching center) via the intermediate nodes 2 and 1; this signal requires only 40 lines since the processor performs a tenfold compression of its 400 incoming circuits. Traffic from all the other nodes is transmitted at the base (unconcentrated) rate to the switching center. By installing the processor at node 5, we have relieved the projected exhaust on edges (2,5), (2,1), and (1,0), i.e., unlike the physical pair relief strategy we do not expand cable capacities in these feeder sections. Observe that we permit traffic flow in either (or both) direction on each edge of the network. For instance, edge (2,5) carries 150 units of unconcentrated traffic (from nodes 2 and 4) from distribution point 2 to the processor located at node 5, while 40 units of concentrated traffic flow in the opposite direction from 5 to 2 (to the switching center). We assume that each unit of traffic (at either transmission rate) requires a single physical line; thus, the 200 available lines in section (2,5) can accomodate both these flows. Also note that the expansion plan shown in Figure 4 involves *backfeed*, i.e., flow that is directed away from the switching center, on section (2,5). Some of the models that we

discuss in Section 4 do not permit such backfeed. Finally, in this example we have implicitly assumed that the switching center can receive signals of varying frequencies from its distribution points. In some practical applications, the designer might specify that all signals entering the switching center must have the same frequency; effectively, this restriction requires that all traffic should undergo equivalent concentration steps. (We can possibly satisfy this requirement by locating a processor at the switching center to process all the unconcentrated traffic.)

The example of Figures 3 and 4 illustrates the two tradeoffs in local access network planning. First, the tradeoff between processors and cable expansion. Installing a processor at node 5 and assigning nodes 2, 4, 5, 8, and 9 to this processor relieves the exhaust in the downstream sections (5,2), (2,1), and (1,0); in our example, the total cost of the traffic processor is lower than the cost of expanding cables along the three sections. We could follow a similar strategy to relieve the exhaust on section (3,1). For instance, locating a processor at node 6 (with a capacity of 120 units) to process node 6's traffic would relieve the 100 units exhaust on edge (3,1). However, we might prefer to expand cable capacities if the cost of a 120-unit processor at node 6 exceeds the cable expansion cost (for 100 additional circuits on section (3,1)). The second tradeoff concerns how many processors to use, and where to locate them. For example, should we locate two processors, one at node 2 and the other at node 5, instead of a single processor at node 5? In the example, locating a single processor at node 5 does not entail any additional cable expansion relative to the two-processor solution. In general, however, the total cable expansion cost might possibly increase as the number of processors decreases. On the other hand, installing fewer, but larger, processors reduces the total processing cost because of economies of scale. The planning model must, therefore, address this tradeoff between exploiting economies of scale in processor costs and avoiding transmission capacity expansion by employing a decentralized processor location strategy.

Our example considered a single processor type, and a single medium that can transport both concentrated and unconcentrated traffic. Additional complexities arise when we consider multiple processor types, multiple processing steps in series, and multiple transmission media. We next present a mathematical programming formulation for this more general problem. We emphasize that this formulation serves only to formalize the problem definition. We will not use it as the basis for any specific solution methods.

3.3 Mathematical Programming Formulation

Our formulation uses two sets of decision variables: y variables to denote processor throughputs at each node, and x variables to represent cable usage. For each node i and processor type m, we define the *processor throughput* variable y_{im} as the throughput (specified in terms of number of incoming lines) of a type m processor(s) located at node i. Similarly, let the *cable usage* variable x_{ijk} represent the number of lines of (sectional) cable type k required on section (i,j) ε E; for continuous cables, let x_{ik} be the number of type k lines connecting node i to the switching center. In turn, the x_{ijk} and x_{ik} variables are decomposed by transmission frequency. Let x_{ijl} denote the number of lines required from i to j (in section (i,j)) to handle communication at transmission rate l; we similarly define the detailed variables x_{il} for rates that require a continuous medium. Observe that our x_{ijl} variables are directed, i.e., we have two variables x_{ijl} and x_{jil} with opposite directions for each edge (i,j) ε E, and every transmission rate l. Thus,

$$x_{ijk} = \sum_{l \varepsilon L(k)} (x_{ijl} + x_{jil}), \tag{1}$$

for all edges (i,j) ε E, and cable types k ε K', where L(k) is the index set of transmission rates requiring (sectional) medium k.

We treat both the x and y variables as continuous variables; the cost functions may reflect discontinuities and non-linearities caused by the discrete nature of processor and cable capacities.

The fundamental constraint in the problem formulation concerns the relationship between incoming and outgoing traffic at each node for each transmission rate. Consider first a transmission rate l that requires sectional cables. Let $M_1(l)$ represent all processor types that have input frequency equal to rate l (i.e., $l_1(m) = l$ for all m ε $M_1(l)$). Similarly, $M_2(l)$ is the set of processor types with output frequency at rate l. (For the lowest rate $l = 1$, the set $M_2(l)$ is empty.) Then, the number of lines emanating from node i carrying rate l transmission equals (i) the number of incoming rate l lines, plus (ii) the number of output lines required for processors located at node i that transmit output at rate l, i.e., for all processor types m ε $M_2(l)$, plus (iii) the demand originating at node i corresponding to service type l,

minus (iv) the number of input lines for processor types in the set $M_1(l)$, accepting input at rate l, located at node i. Mathematically, this relationship can be expressed as follows:

$$\sum_{(j,i)\epsilon E} x_{jil} + \sum_{m\epsilon M2(l)} y_{im}/\rho_m + d_{il} = \sum_{(i,j)\epsilon E} x_{ijl} + \sum_{m\epsilon M1(l)} y_{im}, \qquad (2a)$$

where ρ_m represents the conversion ratio for a type m processor, and d_{il} is the demand for service type l expressed in terms of number of rate l lines. (Since all processor throughputs are expressed in terms of number of input lines, dividing y_{im} by ρ_m gives the number of output lines required for a type m processor that transmits output at rate l.) If rate l requires a continuous rather than sectional medium, the equation changes to:

$$\sum_{m\epsilon M2(l)} y_{im}/\rho_m + d_{il} = x_{il} + \sum_{m\epsilon M1(l)} y_{im}. \qquad (2b)$$

For designing a new network with no existing processing or transmission capacities, the problem formulation becomes:

$$\text{Minimize} \sum_{(i,j)\epsilon E} \sum_{k\epsilon K'} f_{ijk}(x_{ijk}) + \sum_{i\epsilon N} \sum_{k\epsilon K''} f_{ik}(x_{ik}) + \sum_{(i,j)\epsilon E} g_{ij}(x_{ij}) +$$

$$\sum_{i\epsilon N} \sum_{m\epsilon M} h_{im}(y_{im}) \qquad (3)$$

subject to
> constraints (1), (2a), and (2b), and
> non-negativity constraints for all variables,
> for all edges $(i,j) \epsilon E$, cable types $k \epsilon K$, and transmission rates $l \epsilon L$.

Here $x_{ij} = \{x_{ijk}\}$ is the vector of line requirement variables on feeder section (i,j).

To incorporate existing cable and processor capacities, we define an additional set of variables called cable and processor *expansion* variables, denoted as x'_{ijk} (or x'_{ik} for continuous cable types k) and y'_{im}, respectively. We can relate these expansion variables to the original processor throughput and line requirement variables by adding the following constraints to the formulation:

$$x'_{ijk} \geq x_{ijk} - B_{ijk} \qquad \text{for all } (i,j) \, \varepsilon \, E, \, k \, \varepsilon \, K', \qquad (4a)$$

$$x'_{ik} \geq x_{ik} - B_{ik} \qquad \text{for all } i \, \varepsilon \, N, \, k \, \varepsilon \, K'', \text{ and} \qquad (4b)$$

$$y'_{im} \geq y_{im} - A_{im} \qquad \text{for all } i \, \varepsilon \, N, \, m \, \varepsilon \, M. \qquad (5)$$

(Recall that B_{ijk}, B_{ik}, and A_{im} denote, respectively, the existing type k cable capacity on section (i,j), the existing type k (continuous) cable capacity from node i to the switching center, and the existing type m processor capacity at node i.) Correspondingly, we replace x_{ijk}, x_{ik}, and y_{im} in objective function (3) with x'_{ijk}, x'_{ik}, and y'_{im}, respectively. With non-negativity constraints on all the expansion variables, and assuming that all costs are positive, the optimal solution will set

$$x'_{ijk} = \text{Max} \{ 0, x_{ijk} - B_{ijk} \},$$
$$x'_{ik} = \text{Max} \{ 0, x_{ik} - B_{ik} \}, \text{ and}$$
$$y'_{im} = \text{Max} \{ 0, y_{im} - A_{im} \}$$

as desired.

Observe that our formulation models multiple processing steps in series, i.e., traffic originating at a node might possibly be compressed at two or more downstream nodes before reaching the switching center. Also, this formulation permits *bifurcated* routes, i.e., two customers connected to the same distribution point may communicate with the switching center via different routes and processing steps. To prevent this bifurcation, we require a different formulation that distinguishes the traffic originating at different nodes. This latter formulation can also incorporate various proximity restrictions.

Depending on the specific application context, the formulation may contain additional variables and/or constraints. For instance, we can model a processor cost structure that contains fixed and variable components by introducing additional binary variables z_{im} denoting whether ($z_{im} = 1$) or not ($z_{im} = 0$) a new type m processor is installed at node i. The fixed cost of this processor, say H_{im}, serves as the objective function coefficient for variable z_{im}. And, to relate the 'location' variable z_{im} to the capacity expansion variable y'_{im}, we add the forcing constraint

$$y'_{im} \leq A'_{im} z_{im} \, ,$$

where A'_{im} is the maximum permissible capacity of the new processor. Similarly, the formulation may contain additional constraints to model certain policy

restrictions. In Section 4.3 we show that under certain assumptions, we can define all the decision variables in terms of equivalent base rate channels instead of number of lines for different media. This alternative definition enables us to use a network flow representation that conserves flow at each node (unlike the formulation just described).

3.4 Possible Modeling Assumptions

The problem formulation described in Section 3.3 defines a basic framework for all the local access network planning models that we describe in Section 4. Within this framework, different modeling approaches are possible, each characterized by an additional set of assumptions. These assumptions are motivated by three factors:

(i) *model tractability*: With some simplifying assumptions the model becomes more tractable from a computational point of view. For instance, certain network location problems are NP-hard when they are defined over general networks, while they are polynomially solvable for tree networks;

(ii) *uncertainty in technology*: Since the local access network technology is still evolving, we must make some assumptions based on an assessment of the future capabilities and configurations of switching and transmission facilities. The uncertainty about device requirements for future advanced customer services is one such example; and

(ii) *differences in planning practices and expansion policies*: Certain company and country-specific practices and policies give rise to different sets of assumptions. For instance, some local operating companies may emphasize non-bifurcated routing to reduce the burden of managing/rearranging the network, while others may permit bifurcated routing.

Next, we discuss some dimensions that differentiate the various possible modeling approaches.

(a) *New versus Expansion projects*: From a computational point of view, the problem of designing a new network (with no existing capacities) is simpler compared to an expansion planning model where we must account for existing switching and transmission resources. As we showed in section 3.3, expansion planning models require an additional set of *expansion* variables

and associated constraints (constraints (4a), (4b), and (5)) that make the model more difficult to solve.

(b) *Tree versus General networks*: As we mentioned previously, assuming a tree structure for the feeder network reduces the problem complexity, and in some cases makes the model solvable in polynomial time. The solution efficiencies result from the fact that tree networks have a unique path from each distribution point to the switching center; general networks require additional decision variables to determine the route for each distribution point's demand.

(c) *Unidirectional versus Bidirectional flows*: Our example of Section 3.2 illustrated the added dimension when we permit *backfeed*, i.e., traffic movement away from the switching center. Without backfeed (i.e., when all arcs are unidirectional, in the direction of the switching center), a processor that is located at node i can only serve upstream distribution points; this restriction limits the space of solutions that the algorithm must search.

(d) *Processor and Transmission cost functions*: In Section 3.3, we used generic cost functions for processor and transmission capacity expansion. These cost functions can be specialized in various ways. For instance, if all costs are purely variable, and are linear functions of capacity, we can apply a network flow model (possibly with gains) to solve the local network planning problem. When we include a fixed charge for each network enhancement, the problem becomes much more difficult to solve. Similarly, ignoring the joint costs reduces model complexity, as does the assumption that all high frequency transmission media are continuous rather than sectional.

(e) *Routing restrictions*: We have already mentioned one possible distinction in routing strategies: *bifurcated* versus *non-bifurcated* routing. Non-bifurcated routing specifies that all the traffic from each distribution point must follow the same route (i.e., they must use the same feeder sections, the same transmission medium on each section, and the same nodal processors at intermediate distribution points). This policy facilitates network management and maintenance. Another possible routing restriction may

specify that if a node contains a traffic processor, then all traffic entering that node must be processed at that node.

(f) *Multiple versus Single processing steps*: Our formulation of Section 3.3 permits multiple traffic processing steps in sequence. However, existing local access networks rarely employ this progressive traffic compression scheme because it is often not economically viable. If we assume that the traffic from each distribution point can be processed at most (or exactly) once, the number of alternative *homing patterns* (i.e., assignment of traffic from various nodes to processors) to consider decreases significantly, thus reducing model complexity.

These six dimensions illustrate the diversity of models that might apply to local access network planning. In the next section, we outline some possible modeling approaches for the single-period local network planning problem, and differentiate these approaches along the six dimensions. We first review some models from the literature, and subsequently describe two new models.

4. Modeling Approaches for Local Access Network Planning

The telecommunications literature contains several models for contemporary problems in long-distance network design (e.g., Yaged (1971,1973), Zadeh (1973,1974), Ash et al. (1981), Helme and Magnanti (1989), Sinha et al. (1986), Oda et al. (1986)) and inter-office network planning (e.g., Combot and Epstein (1979), Combot and Mason (1980), Combot et al. (1981), Mason (1983, 1984), Doverspike (1986), Sen et al. (1988), Hoang and Lau (1982)). In contrast, models that are specifically tailored for local access network modernization are not widely discussed in the literature. This section describes some modeling approaches for single-period local access network planning. All the models we consider conform to the general framework proposed in Section 3. For each approach, we focus our discussions on the assumptions that differentiate it from other models, and briefly indicate the solution strategy. We do not present any detailed model formulation or algorithmic features. Sections 4.1 reviews existing models, while sections 4.2 and 4.3 cover two alternative approaches. Section 4.1 first considers the general class of models called concentrator location models. We discuss the different versions of this problem found in the literature. A brief discussion of two other models concludes this section. In Section 4.2, we present a network design problem formulation that generalizes several previous concentrator location models. We describe a layered network representation that forms the basis for this model, and indicate some possible solution approaches. Section 4.3 describes a more restrictive model that applies only to tree networks. Because of its simplifying assumptions, this model can be solved efficiently when the problem involves designing new networks.

4.1 Concentrator location and other design models

4.1.1 Centralized Teleprocessing Design

In the 1960's and 1970's centralized teleprocessing systems were quite common, and configuring networks to connect users of the system to the central computer was an important design issue (see Boorstyn and Frank (1977), Chandy and Lo (1973), Chandy and Russell (1972), Direlten and Donaldson (1976), Kershenbaum and Boorstyn (1975), Kershenbaum and Chou (1974), Mirzaian (1985), and McGregor and Shen (1977)). These networks typically consist of many (usually

100 or more) geographically dispersed terminals that are connected to a central computer via communication lines. The central computer provides computational resources, and acts as a switch to connect the terminal to a wider distributed computation network. When the number of terminals is large, and these terminals are located in clusters, using concentrators to combine the communications traffic from several terminals (to increase the utilization of communication lines) becomes a cost-effective interconnection strategy. Certain nodes of the network are preselected as potential concentrator sites, and the set of all available lines (direct connections) are specified.

The teleprocessing network design problem consists of three main components: (1) specifying the number and location of concentrators (*concentrator location*), (2) assigning terminals to either a concentrator or the central processing unit (*terminal assignment*), and (3) determining how to connect every concentrator or central processing unit to its assigned terminals (*terminal layout*). The similarity with the local access network planning problem becomes apparent when we treat terminals as distribution points, and the central computer as the switching center. Ideally, because of their interdependence, we should consider all three teleprocessing design decisions simultaneously in an integrated model. However, the combined model becomes a large-scale integer programming problem which is difficult to solve. Hence, many teleprocessing network design methods proposed in the literature first determine the concentrator location and terminal assignment decisions using a single model (sometimes called the *capacitated concentrator location problem*) that approximates the cost of connecting terminals to concentrators. Subsequently, a terminal layout method may be applied to configure the terminal-to-concentrator networks based on the assignments suggested by the first phase.

Before describing some of the proposed solution approaches for teleprocessing network design, we first discuss the models' underlying assumptions as they relate to the local access network planning context. We focus on the capacitated concentration location problem (CCLP).

Capacitated Concentrator Location Models

This problem selects concentrator locations, and assigns terminals to the selected concentrators in order to minimize the total concentrator and terminal assignment costs, subject to concentrator capacity constraints. CCLP models typically make the following assumptions:

(1) They only apply to the design of *new* networks, i.e., they do not account for existing transmission or concentrator capacities.

(2) CCLP models do not address the problem of configuring the network that connects the concentrators to the central computer. Typically, they assume that each concentrator is directly connected to the central computer. Effectively, the concentrator-to-computer transmission costs become separable by concentrator; hence, these costs can be incorporated directly in the concentrator costs, instead of separately considering transmission costs. Most CCLP models account for a fixed concentrator cost that varies by location; some models also include a variable cost that varies linearly with the concentrator throughput.

(3) As we mentioned previously, the CCLP does not consider the detailed terminal layout decisions either. Instead, it requires user-specified terminal-to-concentrator assignment costs to determine the terminal assignments; these costs may vary by terminal-concentrator pair. Effectively, this cost structure also implies that the terminal-to-concentrator costs are separable, i.e., the cost of connecting a terminal i to concentrator j does not depend on which other terminals are connected to j. Some CCLP models indirectly account for the cost economies when terminals are connected by a spanning tree (Woo and Tang (1973)). Observe that we can enforce proximity restrictions in the CCLP model by setting very high assignment costs for prohibited assignments.

(4) The CCLP assumes that each terminal must be connected to exactly one concentrator, and does not permit bifurcated routing. Most CCLP models also provide for only one level of concentration; an enhanced version of the model, called the *multilevel concentrator location problem* designs a hierarchical structure, where concentrators from one level home on concentrators at the next higher level, and so on.

(5) CCLP models assume a single service type. They also assume a single concentrator type, though multiple concentrators can be modeled by replicating the nodes corresponding to the potential concentrator locations, and associating a different concentrator type with each copy. For each concentrator, the model

assumes a fixed capacity; the total demand for all the terminals that are assigned to a particular concentrator must not exceed this capacity. Some specialized versions of the problem limit only the number of terminals assigned to a concentrator, effectively assuming that all terminals have equal demand. Unlike the local access design problem, the CCLP does not have any provisions for expanding concentrator capacities.

Algorithms for the CCLP belong to three broad classes: adaptations of plant location solution algorithms, heuristic local improvement methods, and clustering techniques.

The CCLP is structurally related to the *plant location* problem, which has been studied extensively in the literature (Cornuejols et al. (1977), Efroymson and Ray (1972), Erlenkotter (1978), Feldman et al. (1966), Kuehn and Hamburger (1963), Sa (1969), Spielberg (1969)). Given a set of potential plant sites and customer locations, the plant location problem seeks the optimal location of plants and assignment of customers to satisfy the customer demands at minimum total plant investment and customer-to-plant transportation costs. Researchers have successfully solved some relatively large-scale uncapacitated plant location problems optimally (Erlenkotter (1978), Geoffrion and Graves (1974)), even though this problem is theoretically intractable. For the CCLP, plants correspond to concentrators and customers to terminals, with the additional restriction on plant capacities. Woo and Tang (1973) propose a CCLP algorithm based on plant location solution methods.

Local improvement procedures for CCLP start with an initial set of concentrator locations and terminal assignments, and attempt to sequentially decrease the total cost by performing myopic changes. The Add heuristic and the Drop heuristic are two common local improvement methods. The Add algorithm (Kuehn and Hamburger (1963)) is a perturbation method that iteratively evaluates the net savings (savings in terminal assignment costs less the cost of an additional concentrator) that accrue by adding each unused site to the current set of concentrator locations. If no site produces net savings, the method terminates. Otherwise, the most cost-effective site is added to the current set, and the method reevaluates net savings for all remaining sites. Conversely, the Drop algorithm (Feldman et al. (1966)) iteratively removes currently selected concentrators until no further cost reduction is possible. The starting solution for the Drop heuristic

locates concentrators at every possible site, and assigns each terminal to its nearest concentrator. Researchers have also proposed combined methods that alternate between Add and Drop phases.

Heuristic methods for the CCLP based on *clustering* concepts have been proposed by McGregor and Shen (1977), Schneider and Zastrow (1982), and Konangi et al. (1984). The first paper deals with the single level concentrator location problem, while the other papers address the multilevel problem. We briefly review the method proposed by Konangi et al. (1984). This method assumes that the concentrator location problem is defined over an Euclidean network, i.e., the terminal-to-concentrator assignment costs are proportional to the Euclidean distance between the two locations. For each concentrator level, the method first clusters the terminals into two groups based on geographical proximity, and locates a concentrator at the center (or the potential site closest to the center) of each cluster. (Observe that locating the concentrator at the center of each cluster is appropriate when concentrator costs do not vary significantly by location.) Clusters are then successively split if savings result from this splitting, and concentrators are relocated at the centers of the new clusters. When cluster splitting does not give any further savings, a cluster merging procedure attempts to further reduce cost by combining previously defined clusters. After completing the clustering and merging steps for one concentrator level, the algorithm considers the next level, treating the concentrators in the current level as the new terminal locations. Computational results are reported for problems with over 200 terminals.

Finally, the literature contains some other *optimization-based* methods for the single-level CCLP. For instance, Pirkul (1986) proposes the following method based on Lagrangian relaxation. This model assumes that concentrators are available in any desired size, subject to a prespecified maximum capacity limitation (that may vary by site); to incorporate the concentrator sizing decision, the model permits both fixed and variable concentrator costs. Like the other CCLP models, this model also incorporates terminal-to-concentrator assignment costs, and does not determine the specific topology of the interconnection network (implicitly assuming direct terminal-to-concentrator and concentrator-to-computer connections). The author formulates this problem as a mixed integer program, and dualizes (i.e., transfers to the objective function using multipliers) the terminal-to-concentrator assignment constraints. The residual subproblems are single constraint 0-1

knapsack problems, one for each concentrator location (the single remaining constraint corresponds to the maximum concentrator capacity restriction). Solving these subproblems gives a lower bound on the cost of the original problem; to obtain good lower bounds, a subgradient optimization method iteratively changes the multipliers. At each iteration, a heuristic procedure uses the Lagrangian subproblem solution to construct a feasible solution, which provides an upper bound. Computational results are reported for problems with up to 100 nodes and 20 concentrator sites; the % gaps between the best upper and lower bounds vary from 0% to 7.7%.

Terminal Layout problem

Given the assignment of terminals to concentrators, the *terminal layout problem* seeks the best network topology connecting each concentrator to its assigned terminals. This model makes the following assumptions:

(1) The original network, containing all available lines, has a general structure; the final topology that is selected must have a tree structure.

(2) Each edge of the network carries a fixed charge; the model does not account for variable edge costs, effectively ignoring cable sizing decisions.

(3) The model only incorporates certain special types of capacity constraints, that apply mostly to multidrop lines. These constraints include: (a) *degree constraints*, which specify an upper limit on the number of incident links at a branching node, or at the concentrator, (b) *order constraints*, which restrict the number of intermediate branching nodes between any terminal and the concentrator, and (c) *load constraints*, which limit the total number of terminals that are connected via a multipoint line.

Observe that by considering the central node as the switching center (instead of a concentrator), and the terminal nodes as concentrators, the terminal layout model applies to the topological design of concentrator-to-switching center interconnections as well.

Several authors have proposed heuristic methods for the terminal layout problem (e.g., Esau and Williams (1966), and Sharma (1983)). The Esau-Williams procedure begins with a star network connecting each terminal directly to the concentrator. The method performs a series of edge interchanges to monotonically decrease costs while satisfying the multipoint line capacity restrictions, and

terminates when no further cost reduction is possible. A special case of the terminal layout problem that has received special attention in the optimization literature is the *capacitated minimal spanning tree problem*. This problem seeks the minimal spanning tree connecting the concentrator to the terminals, subject to degree constraints at the concentrator. Several optimal and optimization-based heuristic algorithms have been proposed for this problem (see, for example, Chandy and Lo (1973), Gavish (1983), Gavish and Altinkemer (1986)).

Finally, some researchers have proposed composite algorithms that combine the CCLP and terminal layout decisions. For instance, Rousset and Cameron (1986) developed the following method for designing private data networks for the French telecommunication system. The procedure assumes that concentrator locations are prespecified. The first phase heuristically assigns terminals to concentrators by first calculating, for each terminal, the difference in assignment costs between the nearest and second nearest concentrators. Terminals are sorted in decreasing order of this difference and assigned, if possible, to their nearest concentrator. Each concentrator has a capacity constraint which must be satisfied in the assignment process. The next module applies the Esau-Williams heuristic to optimize the topology of the subnetwork connecting each concentrator to its assigned terminals using multidrop lines. The final phase attempts to change terminal-to-concentrator assignments in order to reduce the total cost further. The authors report a 15% savings using this method compared to manual designs.

4.1.2 Other models

We now describe two other local access network design models, both of which apply only to the design of new networks, i.e., they do not account for existing cable or processor capacities. The first method is a heuristic proposed by Luna et al. (1987), and the second is a dynamic programming algorithm developed by Helme et al. (1988).

Luna et al. (1987) consider the problem of designing a minimum cost network that connects the switching center to a given set of customers (or distribution points) with known demands. We will refer to this model as the *switching center connection problem*. The distribution points are partitioned into S subsets called service sections; the user also specifies a set of potential concentrator sites within

each service section, and a fixed cost for each site. (The authors do not refer to the nodes as distribution points or concentrator sites; we attach this interpretation in order to relate the model to our local access network planning problem.) Exactly one site must be selected from each service section, and all distribution points within a service section must be served by this site. The original network has a general topology containing all permissible interconnections. Each arc of the network carries a fixed cost (for using that arc) as well as a variable cost that depends on the volume of traffic that is routed on that arc. The switching center connection problem involves (i) selecting one concentrator site from each service section, and (ii) designing a subnetwork (of the given original network) that connects all the selected concentrator sites to the switching center. The objective is to minimize the total (fixed + variable) arc costs plus the concentrator costs. Unlike the CCLP, the switching center connection problem explicitly considers the topological design decisions for connecting concentrators to the switching center. However, the model ignores the interconnections within each service section, i.e., it does not consider the topological design decisions for the subnetwork connecting the distribution points within a service section to the selected concentrator site in that section. Effectively, it assumes that, for each potential concentrator site, the designer has predetermined the distribution point-to-concentrator connections; the cost of this subnetwork can then be incorporated in the concentrator cost. The model does not consider multiple services, multiple concentrator types or different transmission media, and it does not model economies of scale in concentrator and transmission costs.

Observe that proximity restrictions are implicitly incorporated in the choice of service sections. Also, if we interpret the nodes of the network as individual customer locations, the service sections as allocation areas, the potential sites within each section as potential distribution points, and the switching center as a concentrator, then Luna et al.'s model applies to the design of the distribution network for each concentrator, rather than the feeder network serving the switching center.

Luna et al. first formulate the switching center network design problem as a mixed integer program, and propose a heuristic method to solve it. The heuristic starts by constructing the following design: for each service section, select the distribution point that is closest to the switching center; and, find the shortest path

tree (using only the variable arc costs) connecting the switching center to each selected concentrator site. A local improvement procedure then attempts to improve this initial feasible solution. The method iteratively evaluates the cost savings when concentrator sites or arcs are interchanged; profitable interchanges are performed sequentially until no further savings result. Computational results are reported for 3 problems ranging in size from 18 nodes, 54 arcs, and 7 service sections, to 263 nodes, 752 arcs, and 117 service sections. However, the quality of these solutions cannot be evaluated since the method does not generate any lower bounds or alternative heuristic solutions.

Helme et al. (1988) propose a dynamic programming method to solve a problem that is quite similar to our local access network planning model. The method permits multiple processors in series, but assumes a single transmission medium and applies only to the design of new networks. It assumes that the given network has a tree structure, does not permit backfeed, and does not account for economies of scale. Each processsor has an associated fixed cost (that may vary by location); transmission facilities have only variable costs. The method is based on a recursive procedure that exploits the tree structure. For each node of the network, the recursive relationship determines the cost of connecting that node to the switching center, for various possible combinations of downstream processor locations.

Helme et al. also present a mixed integer programming formulation and describe a drop/add heuristic for the general local access network planning problem. The mixed integer formulation applies to general network topologies and multiple processor types in series, permits bidirectional transmission (i.e., backfeed) on links, and incorporates existing capacities as well as fixed and volume-dependent cable and processor costs. Although the mixed integer formulation is very general, it can be solved optimally only for very small problem sizes. The authors propose a greedy drop/add heuristic that assumes linear cable costs and permits backfeed. The drop heuristic starts with an initial design containing all processor types at each node, and successively eliminates processors to reduce the total cost. For any given processor configuration, the method applies a shortest path algorithm to compute the total connection cost; this algorithm exploits the linear cable expansion cost structure.

4.2 Network Design Model

Fixed-charge network design problems arise in a variety of distribution planning, manufacturing and telecommunications contexts. Given the demand between various origin-destination pairs, and fixed and variable costs for each arc of a network, the (fixed charge) network design problem involves selecting a subset of arcs, and routing the various commodities over the selected arcs in order to minimize the total fixed plus variable arc costs. The capacitated version of this problem accounts for additional arc capacity constraints. The network design problem generalizes several well-known optimization models including the plant location, shortest path, Steiner tree, traveling salesman, and minimal spanning tree problems. Magnanti and Wong (1981) describe various applications and solution methods for the network design model.

In this section, we demonstrate how to transform the local access network planning problem under certain assumptions to a network design model. The model is defined over a layered network representation that contains a separate layer for each transmission rate. We first outline the model's assumptions before describing the layered network.

Our network design model makes two main simplifying assumptions. The first assumption concerns the transmission and processor cost structures. Our model ignores all joint costs between various transmission media. Furthermore, it associates a different cable type with each transmission rate, effectively ignoring possible cost interactions when the same physical medium (which may now correspond to different cable types) carries different transmission rates. It also assumes that all processor and cable installation/expansion costs are piecewise linear, consisting of (possibly) fixed and variable components.

As before, we index the cable types from 1 to $|K|$ in increasing order of transmission rate; our assumption regarding a unique transmission rate for each cable type implies a one to one correspondence between the set of cable types $K = \{1,2,...,|K|\}$ and the set of available transmission rates $L = \{1,2,...,|L|\}$.

The model's second assumption concerns the conversion ratios for different processor types. Recall from our discussions in Section 3.1 that, in general, each processor type m requires a certain input rate (or frequency) denoted as $l_1(m)$, transmits output at a higher rate $l_2(m)$, and, has a specified conversion ratio ρ_m (defined as the ratio of input to output lines). Our model assumes that the conversion ratios for different processor types are *compatible* in the following sense. Consider three processor types labeled 1, 2, and 3, and suppose the

input rate of type 1 = input rate of type 3, i.e., $l_1(1) = l_1(3) = l_a$

output rate of type 1 = input rate of type 2, i.e., $l_2(1) = l_1(2) = l_b$, and

output rate of type 2 = output rate of type 3, i.e., $l_2(2) = l_2(3) = l_c$.

Effectively, we can convert traffic from rate l_a to rate l_c either by employing a type 1 and type 2 processor in series, or by using a single type 3 processor. The conversion ratio compatibility assumption requires that the three conversion ratios must satisfy the following equation:

$$\rho_3 = \rho_1 * \rho_2.$$

In other words, the messages on x lines in the rate l_a medium always require exactly $(x + \rho_3)$ lines in the rate l_c medium, regardless of whether the compression was achieved using a type 1 and type 2 processor in tandem, or a single type 3 processor. Effectively, this assumption permits us to associate a single *conversion factor*, call it δ_k, with each cable type k. We define this factor as the number of channels (or circuits or lines) at the *base rate* (the lowest transmission rate, i.e., corresponding to the index $l = 1$) that each type k line can accommodate. In the above example, if l_a is the base rate, then the conversion factor for the cable type corresponding to rate l_c is ρ_3. In turn, this single conversion rate for each cable type enables us to measure all the traffic in every medium in terms of the number of equivalent base rate channels (rather than the number of lines of the corresponding medium). This traffic measurement unit preserves conservation of flow at each node, as we describe later.

Apart from these two assumptions, the network design model incorporates all the other features of the general problem described in Section 3.2. In particular, it can handle general network topologies, multiple service types (as long as these do not have unique processing requirements), sectional and continuous cable types, economies of scale in processor and transmission cost functions, and existing transmission and processor capacities. It also permits backfeed and bifurcated routing. If the cost functions are piecewise linear and concave, and if the network

does not contain any existing capacities, the model reduces to an uncapacitated network design problem. Existing resources and non-concave cost functions introduce arc capacities. We first describe the layered network representation for the uncapacitated model with no existing processor and transmission capacities, and with a fixed plus linear cost structure for each transmission and processing facility. Subsequently, we discuss extensions to model piecewise linear, concave functions, existing capacities, and non-concave costs. Our layered network builds upon similar representations previously proposed by Helme et al. (1988) and Sen et al. (1989).

Layered Network Representation

The single-layer network representation of the local access network planning problem described in Section 3.2 is inadequate since it does not establish the relationship between the different processor types and their respective input and output cable type requirements. The layered network shown in Figure 5 provides a more natural problem representation. In this network, each layer corresponds to a different transmission rate l; by our assumptions, each layer also corresponds to a unique transmission medium $k = l$, and carries a conversion factor δ_k that specifies the number of base rate channels that each medium k line can accommodate. For convenience, we index the layers in increasing order of traffic compression, with layer 1 (corresponding to, say, twisted wire pairs) carrying traffic at the base (or lowest) rate, and layer $|L|$ (corresponding to fiber optic transmission, for instance) carrying the highest possible rate.

The nodes of the original network are replicated in each layer. We denote the copy of node i in layer l as (i,l). Layer l contains an edge between node i and node j if transmission medium l can be installed in the feeder section from i to j (assuming this feeder section exists or is permitted in the original problem definition). We will refer to edges within each layer as *transmission edges*; the transmission edge connecting nodes i and j in layer l is denoted as (i,j,l). Flows along the transmission edges in the lth layer will represent physical transmission on cable type l. The representation of Figure 5 assumes, for simplicity, that every transmission medium is sectional (versus continuous), and can be installed in each feeder section; thus, each layer contains a replication of the original (given) network topology.

The edges connecting two different layers represent traffic processors; hence, we will refer to them as *processor edges*. The processor edge from node i in layer l' to node i in layer l'' (with $l'' > l'$), denoted as (i,l',l''), represents a traffic processor at node i that compresses layer l' traffic to layer l'' traffic. The flow through this edge must be less than or equal to the processor's capacity. Observe that, unlike the transmission edges, the processor edges are directed from lower indexed layers to higher indexed layers since we only permit traffic processing from lower to higher rates.

We have two possible units for measuring the flow on each edge in the layered network: either in terms of the number of physical lines required in the corresponding medium (as in the formulation of Section 3.3), or in terms of a common traffic unit, namely, the number of equivalent base rate channels. Using the number of physical lines as the unit of flow measurement destroys the conservation of flow along processor edges since the number of incoming lines exceeds the number of outgoing lines from a traffic processor. We, therefore, use the common traffic unit to measure flows in all layers. Thus, if an edge in layer l of the layered network carries a flow of x units, the number of physical lines required along this feeder section is x/δ_l, where δ_l is the conversion factor corresponding to cable type l. Similarly, we measure the capacities and throughputs of the nodal processors in terms of the common traffic unit.

To complete the layered network description, we must associate demands and supplies with the nodes of the network, and costs and capacities with its edges. We treat each node (except the nodes corresponding to the switching center) in every layer as a demand node. The demand d'_{il} for node (i,l) in layer l, which is also expressed in terms of the common traffic unit (number of equivalent base rate channels), corresponds to the projected traffic at node i for a service that originates at rate l, and requires this minimum transmission frequency. (Previously, we defined demand d_{il} in terms of the number of rate l <u>lines</u>. Our new definition of demand implies that $d'_{il} = \delta_l * d_{il}$.) The switching center node in layer L, denoted as node $(0,L)$, acts as the supply node with total supply equal to the sum of all demands in layers 1 through L. The switching center nodes in all other layers are transshipment nodes.

The transmission and processing costs are represented as edge cost functions in the network. For the moment, let us assume that the network does not have any existing transmission or processing capacity, and that the expansion cost functions are fixed plus linear costs. In particular, let $H_{il'l''}$ be the fixed cost of a processor located at node i that converts layer l' signals to layer l'' signals, and let $v_{il'l''}$ be its per unit cost. Thus, for a processor with capacity of x units, the total cost is $H_{il'l''}$ + $v_{il'l''}*X$. We associate this fixed and variable cost with the processor edge (i,l',l'') connecting layers l' and l''. Similarly, let F_{ijl} and c_{ijl} represent, respectively, the fixed and per unit costs for a type l link from node i to node j. (Note that a per unit cost of c_{ijl} implies that each additional medium l <u>line</u> from i to j costs δ_l*c_{ijl}.) These two parameters define the fixed and variable costs for the transmission edge (i,j,l). Finally, the cost parameters for the inter-layer edges connecting the switching center nodes $(0,l')$ and $(0,l'')$, with $l' < l''$, depend on our assumption regarding permissible transmission rates for traffic entering the switching center. In particular, if we permit multiple signal rates entering the switching center, then each of these edges carries zero costs. Otherwise (if all entering traffic must be at the same rate), the 'processor' edge $(0,l',l'')$ carries the fixed and variable costs corresponding to a l'-to-l'' processor located at the switching center.

With this set of model parameters, the uncapacitated network design solution that satisfies all demands at minimum total fixed plus flow costs corresponds to the optimal local access network plan. In the optimal network design solution, a flow of x_{ijl} units along transmission edge (i,j,l) implies that the number of medium l lines to install in feeder section (i,j) is x_{ijl}/δ_l. Similarly, the flow on processor edge (i,l',l'') divided by $\delta_{l'}$ gives the capacity (in terms of number of input lines) of a processor at node i that transforms layer l' input signals to layer l'' output signals. Observe that, since the transmission edges are bi-directional, the optimal solution might involve backfeed. Also, the model permits multiple processes in series.

This network design model can be enhanced in various ways. For instance, suppose medium l is a continuous rather than sectional medium. In this case, instead of replicating the original network in layer l, we have a star network in this layer, i.e., each node (i,l) is directly connected to the switching center node $(0,l)$ in that layer. The fixed and variable costs of this edge represent the fixed and per unit costs of the continuous type l medium connecting node i to the switching center.

We can also model economies of scale in processor and transmission costs if these economies are adequately represented by piecewise linear concave cost functions as shown in Figure 6. Suppose the function shown in this figure describes the cost of installing a type l cable on feeder section (i,j). In general, this cost function contains R breakpoints. Breakpoint r occurs at a capacity of B_r, and the slope of the cost function decreases from c_r to c_{r+1} at this point. Let F_r and F_{r+1} denote the y-intercepts of the two line segments that define breakpoint r. We can incorporate this cost function in the network design model by introducing R parallel arcs between nodes i and j in layer l. The rth parallel arc carries a fixed charge of F_r and a variable cost of c_r. Because the overall transmission cost function is concave, the optimal solution will automatically satisfy the range constraint (B_r to B_{r+1}) for the rth line segment without explicit capacity constraints on the rth edge of the enhanced network, i.e., if the optimal local access network solution entails installing a capacity of x units, with $B_r < x \leq B_{r+1}$, between nodes i and j, the network design solution will route all x units on the rth edge since this edge minimizes total cost, among all parallel edges, for x units of flow. We can similarly model piecewise linear, concave processor cost functions by introducing parallel processor edges. Finally, the single commodity network design formulation that we have just described does not readily accommodate proximity restrictions. To incorporate these restrictions, we require a multicommodity version that separately identifies the traffic originating at each node.

Certain properties of the optimal uncapacitated network design solution (with concave edge cost functions and no existing capacities) have special significance for local access network planning. For instance, we can show that, at each node (i,l), the optimal network design solution either processes all incoming traffic or routes all the traffic to another node on the same level, but not both. Thus, we cannot have both type l traffic leaving node i, and a l-to-l' processor located at node i. In particular, consider the route for, say, layer 1 traffic originating at node i. Let node j be the first node on this route with a 1-to-l processor. Then, the traffic originating at node i must necessarily undergo this processing step. Furthermore, all layer 1 traffic originating at intermediate nodes (on the path from i to j) must also follow the same route as node i's traffic.

The uncapacitated network design model applies to local access design problems with no existing capacities. When the network contains existing processor

and transmission capacities, or if the cost functions are piecewise linear but non-concave, we must add explicit capacities on the edges of the layered network. In particular, suppose the network already contains x type l lines connecting nodes i and j. To represent this capacity we add a parallel arc connecting nodes (i,l) and (j,l) (in layer l); this arc carries zero fixed and variable costs, but has a capacity of $x^*\rho_l$ (basic traffic) units. Similarly, suppose expanding the type l transmission facilities in section (i,j) entails a piecewise linear, convex cost as shown in Figure 7a. Figure 7b shows the equivalent network representation with parallel arcs whose arc fixed costs, variable costs, and capacities are as shown in the figure. Figure 8a shows a more general cost structure, and Figure 8b gives its representation.

In summary, the network design formulation for the local access network planning problem is very versatile. Its assumptions are less restrictive than previous local access network models, and indeed it generalizes many of the previous models. Next, we briefly outline solution methods for the network design problem.

Solution Methods for the Network Design Problem

The network design problem and several of its variants are known to be NP-hard (Johnson, Lenstra and Rinnooy Kan (1978)). Several authors have proposed heuristic and optimal methods for solving the problem (Hoang (1973), Boyce et al. (1973), Billheimer and Gray (1973), Boffey and Hinxman (1979), Dionne and Florian (1979)). Balakrishnan et al. (1989) propose a dual ascent method that generates provably near-optimal solutions to the uncapacitated network design problem. The method essentially involves approximately solving the dual of the linear programming relaxation for the network design integer programming formulation. The dual solution generates a starting design for a local improvement heuristic and also provides a lower bound which can be used to verify the quality of the heuristic solutions. This method generalizes several previous algorithms for special cases of the network design problem, including the Steiner tree and plant location problems. The approach was successfully tested on several randomly generated problems containing upto 45 nodes and 595 arcs.

Capacitated network design problems are much harder to solve compared to the uncapacitated version. One possible solution strategy consists of dualizing the arc capacity constraints (i.e., multiplying the capacity constraints with Lagrange multipliers, and adding these multiples to the objective function). The resulting subproblem is an uncapacitated network design problem which can be solved at least approximately using, say, the dual ascent algorithm. By iteratively modifying the Lagrange multipliers using a method such as subgradient optimization (see, for example, Fisher (1981)), we can possibly generate good heuristic solutions and lower bounds for the original capacitated problem. However, previous experience with this approach for the plant location and other related models suggests that the gaps between the upper and lower bounds are likely to be significantly larger for capacitated problems relative to the gaps for uncapacitated problems. Thus, we expect local access planning problems with existing capacities and non-concave cost functions to be computationally more difficult.

One of the main limitations of the network design model is its size. In particular, the number of nodes and arcs in the layered network grows very rapidly with the number of different transmission rates, processor types, and distribution points. To model a problem involving 20 distribution points with all possible connections, 5 processor types, and 3 transmission media, we require a network with 63 nodes, and over 600 arcs. These problem dimensions probably represent the largest size that current optimization-based network design algorithms can solve within a reasonable amount of computational time. In the next section we describe an alternative specialized model that is more tractable since it assumes a tree network, and restricts the assignment of distribution points to concentrators.

4.3 Tree Covering Model

In this section we describe a special case of the local access network planning problem, which we call the *tree covering model*, that is solvable in polynomial time when the network does not contain any existing capacities. The model assumes that the given network defining the permissible interconnections is a tree network. It also makes some additional assumptions regarding the cost structure and routing policy. It permits backfeed, and can incorporate economies of scale. We first

- 48 -

describe the model as it applies to the design of new networks, and subsequently describe an enhancement to account for existing capacities.

First, we introduce some terminology. We say that a node i (i.e., distribution point) *homes* on another node j if the traffic from node i is processed at node j. Node i homes on the switching center (node 0) if its traffic is not processed at any intermediate node. The tree covering model makes the following assumptions:

(1) The original network has a tree structure, rooted at the switching center. This assumption implies that a unique path connects each distribution point to the switching center.

(2) The model permits at most one level of traffic processing, and assumes a single service type. For simplicity, we assume that traffic can arrive at different frequencies at the switching center.

(3) *Contiguity Assumption*: The model assumes that if a node i homes on node j, then all intermediate nodes lying on the (unique) path between nodes i and j also home on node j. In particular, node j must home on itself if it contains a processor. We refer to this routing restriction as the contiguity assumption since the set of all nodes homing on a particular processor induces a single contiguous or connected subgraph of the original network.

(4) The model does not permit bifurcated routing, i.e., all the traffic originating at a particular node must follow the same route (i.e., they must use the same links, and undergo processing at the same node) to the switching center.

(5) The model permits multiple processor types and transmission media. However, it ignores any joint costs between media, and assumes that all high frequency media belong to the continuous (or umbilical) type. This assumption essentially permits us to include in the processor cost all transmission costs for traffic emanating from each traffic processor. The base rate medium (say, twisted wire pairs), which we will refer to as *cables*, is assumed to be sectional.

(6) Each processor type is assumed to have a fixed plus variable cost structure (including the transmission cost for the processor's output) that may vary by

location. Similarly, cable installation and expansion entails a fixed and variable cost that varies by section. Like the previous the network design model, the tree covering model can also accommodate piecewise linear concave cost functions.

(7) The model can also account for additional *homing costs*, one for each node pair <i,j>, incurred when node i homes on a processor located at node j. For instance, by selectively setting these homing costs to a high value we can prohibit homing patterns that violate proximity restrictions.

This set of assumptions permits us to transform the local access network planning task into a problem of covering the original tree by subtrees. Consider a local access network solution in which node j contains a processor. Let N(j) be the set of nodes that home on this processor, and let T(j) be the subgraph induced by this node subset (i.e., T(j) contains edge (p,q) of the original tree network if both nodes p and q belong to the node subset N(j)). Our contiguity assumption implies that T(j) must be a single connected component, i.e., it must be a subtree of the original tree. Thus, the union of the induced subtrees corresponding to each processor must span all the nodes of the network. (By convention, the switching center always contains a processor.) Conversely, suppose we are given a subtree T that must be served by a processor located at one of the nodes in the subtree. For each potential processor location, we can calculate the exact flows, and hence the exact value of the cable expansion costs for all the edges belonging to this subtree. The processor costs are also known since this processor must serve the sum of the demands for all nodes in the subtree. Consequently, we can easily calculate the total transmission plus processing cost of serving all the nodes in subtree T from each node i ϵ T. The node i that minimizes total costs is the best processor location for this subtree.

These properties enable us to solve the tree covering model (without existing cable and processor capacities) very efficiently using a dynamic programming algorithm based on a method developed by Barany et al. (1986) for optimally covering a tree with subtrees. This method is also closely related to the p-median algorithm discussed by Kariv and Hakimi (1979). The algorithm starts from the leaves of the original tree, and recursively builds the covering solution for successively larger subtrees. Balakrishnan et al. (1989) describe this method in greater detail. Next we outline a different method, using a shortest path algorithm, to solve the special case when the given network is a line network.

A *line network* consists of a simpe path connecting two end nodes, one of which is the switching center node, say, node 0. Without loss of generality, assume that the nodes are indexed sequentially from 0 to n so that the line network only contains (undirected) edges of the form (i-1,i), for i = 1,2,...,n. Figure 9 shows this structure. Suppose we want to locate p processors on this network, including the processor at the switching center. Then, by our contiguity assumption, the nodes served by each processor induce a line segment, and the union of all p line segments cover all the nodes of the network. Thus, the local access network planning problem for this special case reduces to the problem of determining the number of processors (p) to locate, and the optimal partition of the original line network into p segments. We can formulate this problem as a shortest path problem in the following way. Consider a line segment from node i to node j (inclusive), with j > i. As mentioned previously, we can easily determine the optimal total cost of serving all the nodes in this segment by enumerating all potential processor locations between i and j. For instance, consider a potential location k, and suppose k > i+1. Then arc (i,i+1) must carry node i's demand d_i, arc (i+1,i+2) must carry the cumulative demand for nodes i and (i+1), i.e., $(d_i + d_{i+1})$ and so on. Thus, we can determine the flow on each edge of the i-to-j line segment. Also, the total processor throughput is the sum of the demands for all nodes from i to j. Using this information, we can determine the total processor plus cable cost for serving all the nodes between i and j using a processor that is located at node k. Let k_{ij} be the best processor location for serving line segment i-to-j, and let c_{ij} be the corresponding optimal cost.

To find the best partition of the original network, we construct a shortest path network defined over the (n+1) nodes. For every i ≤ j, this network contains a directed arc from j to (i-1). The cost of this arc is set equal to c_{ij}, the optimal cost of serving all nodes between i and j (inclusive). Every path from node n to node 0 then defines a partition of the line network. In particular, including arc (i-1,j) in the n-to-0 path corresponds to selecting the line segment from node i to j as one element of the partition. Consequently, the shortest n-to-0 path defines the optimal partition of the original network, and hence identifies the optimal local access network configuration. Further simplifications are possible for computing the shortest path arc lengths c_{ij}. Also, in this special case of line networks, the algorithm can easily accomodate existing cable and processor capacities.

For general tree networks, incorporating existing cable and processor capacities is not as easy. Balakrishnan et al. (1989) describe a Lagrangian-relaxation approach that first formulates the local access network planning problem as a mixed integer program, and dualizes the capacity constraints. The resulting subproblem is an uncapacitated local access network planning problem that can be solved efficiently using the dynamic programming procedure mentioned previously. This solution gives a lower bound on the total cost of the original problem. The Lagrangian subproblem solution method is embedded in an iterative procedure that modifies the Lagrange multipliers in order to improve the lower bound. The subproblem solution at each iteration also identifies a feasible network expansion plan, which can be improved heuristically to generate good upper bounds. Balakrishnan et al. describe various formulation and algorithmic enhancements to significantly improve the method's performance, and report computational results based on some actual test networks.

In summary, this section has described various models for local access network planning. We have seen the diverse range of possible assumptions, with each combination of assumptions defining a separate model. The fixed-charge network design model is very comprehensive, and recent advances in network design algorithms make this modeling approach computationally feasible for medium-sized problems in the local access network planning context, especially for designing new networks. On the other hand, the general network design model does not exploit any special structure that specific application contexts might possess. For instance, in order to simplify the task of managing the network, some local telephone companies might adopt policies similar to the contiguity assumption. Similarly, if each concentrator requires an umbilical connection to the central office, ignoring shared media costs and assuming a continuous medium for concentrated traffic might be appropriate, especially if these enhanced cables can be installed in existing ducts. Making these simplifying assumptions enables us to use specialized algorithms, thus increasing the range of problem sizes that can be solved.

5. Concluding Remarks

This paper has attempted to trace the evolution of local access network technology in public telecommunication networks as it relates to economic models for design and planning. These models are becoming increasingly important because of rising demand for a variety of services resulting from the introduction of ISDN standards, installation of digital and fiber optic technology, and mounting competitive pressures. The traditional local network planning tools are inadequate in the current environment because digitization and the introduction of electronics within the feeder network has created new ways to respond to increasing demand for telecommunication services.

We described the local access network technology in some detail in order to illustrate the complexity of the planning problem. For our review of planning models, we focused on the static (or single period) problem. In describing the various models, we emphasized the differences in their assumptions, and briefly outlined solution methods. As our discussion of modeling approaches suggests, the general area of local access network planning continues to provide several challenging opportunities for modeling and algorithmic development, particularly for the multiperiod and multiple service contexts. The new developments in telecommunication standards and technologies should further stimulate the development of new modeling approaches.

Some of the static models that we discussed can possibly be extended to a multiperiod framework. For instance, we might employ a decomposition method such as Lagrangian relaxation to decompose the multiperiod problem into several single period problems, which can then be solved using one of the single-period methods. In this scheme, the Lagrangian multipliers corresponding to a time period t might represent the 'price' that we are willing to pay to establish excess transmission and switching resources at time t for use in future periods. Thus, the pricing mechanism accounts for the temporal coupling of plans by acting as an incentive to exploit economies of scale. An alternative use of static models for multiperiod planning is to generate a final target network; we might then apply a different model to plan the evolution, over the multiple time periods of the planning horizon, from the current network to the target network. Shulman and Vachani (1988) propose a related approach.

The models that we discussed did not include any special representation for fiber optic facilities partly because their current economic implications are comparable to those of other electronic traffic processing devices and high frequency media. Future developments and implementation of fiber optics in the local loop might necessitate other distinctions in modeling fiber optic facilities.

The continuing evolution of local access network technology and ever increasing efforts to formulate new ways of utilizing this technology create a number of exciting and challenging future research directions. One interesting technological development is the possibility of installing remote switches and other 'intelligent' hardware in the feeder and distribution networks. These devices can perform a number of switching center functions, and in many applications customers would only need to communicate with a nearby remote switch instead of connecting all the way through the switching center. This strategy of using remote switches would reduce the overall traffic in the feeder network, and thus reduce the need for additional cables or processors. Strategies for the proper deployment of these remote switches is an interesting topic for future exploration.

The enormous bandwidth that fiber optic networks provide creates intriguing opportunities for developing new services for households such as video programming on demand, interactive shopping services and home telemetry. With new services such as home telemetry, customers might become much more dependent on their local telecommunication system and any disruption in service would be very undesireable, perhaps, comparable to a power blackout. Thus, reliability issues should assume a much greater importance in planning for future networks. Because of economic considerations in minimizing the number of links, the most common current local network design is a tree configuration. The disadvantage of this design is that any single link failure will disconnect the network. New research is needed to design local access networks that can offer more reliability and resistance to failure (see, for example, Monma and Shallcross (1986)). Topologies such as ring networks (which provide two paths between every pair of nodes) may become more common in future local telecommunication systems.

References

Anderson, P. (1988) "Fiber optics for loop applications: A techno-economical analysis", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 213-216.

Ash, G., R. Cardwell, and R. Murray (1981) "Design and optimization of networks with dynamix routing", *Bell System Technical Journal*, **60**, pp. 1787-1820.

AT&T Bell Laboratories (1986) *Engineering and Operations in the Bell System*, Murray Hill, New Jersey.

Bahl, L. R., and D. T. Tang (1972) "Optimization of concentrator locations in teleprocessing networks", *Proceedings of the Symposium on Computer Communication Networks and Teletraffic*.

Balakrishnan, A., T. L. Magnanti, and R. T. Wong (1989a) "A dual ascent procedure for large-scale uncapacitated network design", to appear in *Operations Research*.

Balakrishnan, A., T. L. Magnanti, and R. T. Wong (1989b) "A decomposition algorithm for local access telecommunications planning", in preparation.

Barany, I., J. Edmonds, and L. A. Wolsey (1986) "Packing and covering a tree by subtrees", *Combinatorica*, **6**, pp. 221-233.

Billheimer, J., and P. Gray (1973) "Network design with fixed and variable cost elements", *Transportation Science*, **7**, pp. 49-74.

Boffey, T. B., and A. I. Hinxman (1979) "Solving for optimal network problem", *European Journal of Operational Research*, **3**, pp. 386-393.

Boorstyn, R. B., and H. Frank (1977) "Large-scale network topological optimization", *IEEE Transactions on Communication*, **COM-25**, pp. 29-47.

Boyce, D. E., A. Farhi, and R. Weischedel (1973) "Optimal network problem: A branch-and-bound algorithm", *Enviromental Planning*, **5**, pp. 519-533.

Bulcha, B., L. E. Kodrich, D. B. Luber, W. J. Mitchell, M. A. Schwartz, and F. N. Woomer (1982) "Feeder planning methods for digital loop carrier", *The Bell System Technical Journal*, **61**, No. 9, pp. 2129-2141.

Campbell, L. H. (1988) "The evolution from narrowband to broadband customer access", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 144-148.

Carse, G. D. (1986) "New and future technologies in the local telephone network: The Victoria system", *Proceedings of the IEEE 1986 International Communications Conference*, pp. 410-412.

Chandy, K. M., and T. Lo (1973) "The capacitated minimum spanning tree", *Networks*, 3, pp. 173-182.

Chandy, K. M., and R. A. Russell (1972) "The design of multipoint linkages in a teleprocessing tree network", *IEEE Transactions on Computers*, C-21, pp. 1062-1066.

Ciesielka, A. J., and D. C. Douglas (1980) "Electronics in the suburban and light urban loop networks", *The Bell System Technical Journal*, 59, No. 3, pp. 417-439.

Ciesielka, A. J., and N. G. Long (1980) "New technology for loops - A plan for the '80's", *IEEE Transactions on Communications*, COM-28, No. 7, pp. 923-930.

Coathup, L., J.-P. Poirier, D. Poirier, and D. Kahn (1988) "Fiber to the home - Technology and architecture drives", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 281-285.

Combot, J.-P., and N. Epstein (1979) "The introduction of digital switching to the local network", *IEEE Transactions on Communications*, COM-27, No. 7, pp. 1056-1064.

Combot, J.-P., and L. G. Mason (1979) "Optimal switch-concentrator location for a telephone network", *IEEE Transactions on Communications*, COM-27, No. 1, pp. 36-43.

Combot, J.-P., M. S. C. Tsui, and R. Weihmayer (1981) "Optimal digital switching introduction into the local network", *IEEE Transactions on Communications*, COM-29, No. 10, pp. 1446-1454.

Cournejols, G., M. L. Fisher, and G. L. Nemhauser (1977) "Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms", *Management Science*, 23, pp. 789-810.

Dettmer, R. (1985) "The integrated services digital network: Bringing home the bits", *Electronics and Power*, January 1985, pp. 25-28.

Dionne, R., and M. Florian (1979) "Exact and approximate algorithms for optimal network design", *Networks*, 9, pp. 37-59.

Direlten, H., and R. W. Donaldson (1976) "Topological design of teleprocessing networks using linear regression clustering", *IEEE Transactions on Communications*, COM-24, pp. 1152-1159.

Doverspike, R. D. (1986) "Multiplex bundler algorithms for the Network Planning System", *Proceedings of the Third International Network Planning Symposium* (Networks '86), pp. 29-32.

Efroymson, M. A., and T. L. Ray (1972) "A branch-and-bound algorithm for plant location", *Operations Research*, 20, pp. 94-108.

Elken, T. R. (1980) "The application of mathematical programming to loop feeder allocation", *The Bell System Technical Journal*, 59, No. 4, pp. 479-500.

Ensdorf, W. E., M. L. Keller, and C. R. Kowal (1988) "Economic considerations of fiber in the loop plant", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 291-296.

Erlenkotter, D. (1978) "A dual-based procedure for uncapacitated facility location", *Operations Research*, 26, pp. 992-1009.

Esau, L. R., and K. C. Williams (1966) "On teleprocessing system design: Part II", *IBM Systems Journal*, 5, pp. 142-147.

Feldman, E., F. A. Lehrer, and T. L. Ray (1966) "Warehouse location under continuous economies of scale", *Management Science*, 12, pp. 670-684.

Fisher, M. L. (1981) "The Lagrangian relaxation method mfor solving integer programming problems", *Management Science*, 27, pp. 1-18.

Forbes (1986), Novemer 3 issue, pg. 211.

Fortune (1988) "ISDN: The new telephone network", November 1988.

Fortune (1989) "The coming battle", February 13, 1989, pp. 104-107.

Freidenfelds, J., and C. D. McLaughlin (1979) "A heuristic branch-and-bound algorithm for telephone feeder capacity expansion", *Operations Research*, 27, pp. 567-582.

Garbanati, L. F., and J. R. Palladino (1988) "Fiber optics to the subscriber: Objectives, costs, and technology assessment", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 286-290.

Gavish, B. (1983) "Formulations and algorithms for the capacitated minimal directed tree", *Journal of the ACM*, 30, pp. 118-132.

Gavish, B., and K. Altinkemer (1986) "Parallel savings heuristic for the topological design of local access tree networks", *Proceedings of IEEE INFOCOM '86*, pp. 130-139.

Geoffrion, A. M., and G. Graves (1974) "Multicommodity distribution system design by Benders decomposition", *Management Science*, **5**, pp. 822-844.

Gibson, A. E., and D. B. Luber (1980) "Critical section methods for loop plant allocation", *The Bell System Technical Journal*, **59**, No. 1, pp. 99-117.

Griffiths, J. W. (ed.) (1986) *Local Telecommunications II*

Helme, M. P., C. Jack, and A. Shulman (1988) "Planning for new services in the local loop", *Proceedings of the International Telecommunication Conference*, **12**, pp. 5.2B.1.1-5.2B.1.12.

Helme, M. P., and T. L. Magnanti (1989) "Designing satellite communication networks using zero-one quadratic programming", to appear in *Networks*.

Hoang, H. H., and H. T. Lau (1984) "Optimizing the evolution of digital switching in a local telephone network", *Proceedings of the IEEE International Conference on Communications*, pp. 179-184.

Johnson, D. S., J. K. Lenstra, and A. H. G. Rinnooy Kan (1978) "The complexity of the network design problem", *Networks*, **8**, pp. 279-285.

Kariv, O., and S. L. Hakimi (1979) "An algorithmic approach to network location problems II: The p-medians", *SIAM Journal of Applied Mathematics*, **27**, pp. 539-560.

Kershenbaum, A., and Boorstyn (1975) "Centralized teleprocessing network design", *Proceedings of the National Telecommunication Conference*, pp. 27.11-27.14.

Kershenbaum, A., and W. Chou (1974) "A unified algorithm for designing multidrop teleprocessing networks", *IEEE Transactions on Communications*, **COM-22**, pp. 1762-1772.

Kershenbaum, A., and S. Peng (1986) "Neighbor finding algorithms for CMST calculation", *Proceedings of IEEE INFOCOM '86*, pp. 140-147.

Konangi, V. K., T. Aidja, and C. R. Dhas (1984) "On the multilevel concentrator location problem for local access networks", *Proceedings of IEEE Globecom '84*, pp. 912-915.

Koontz, W. L. G. (1980) "Economic evaluation of loop feeder relief alternatives", *The Bell System Technical Journal*, **59**, No. 3, pp. 277-293.

Kopp, P. (1986) "A family of solution procedures for the routing problem in medium-term planning of mixed analog-digital transmission networks", *Proceedings of the Third International Networks Symposium* (Networks '86), pp. 138-142.

Kostas, D. J. (1984) "Transition to ISDN - An overview", *IEEE Communications Magazine*, **22**, No. 1, January 1984, pp. 11-17.

Kuehn, A. A., and M. J. Hamburger (1963) "A heuristic program for locating warehouses", *Management Science*, **9**, pp. 643-666.

Lavin, J. S. (1987) "Optimally rehoming local serving offices to a new point of presence", *AT&T Technical Journal*, **66**, pp. 50-54.

Luna, H. P. L., N. Ziviani, and R. H. B. Cabral (1987) "The telephonic switching centre network problem: Formalization and computational experience", *Discrete Applied Mathematics*, **18**, pp. 199-210.

Magnanti, T. L., and R. T. Wong (1984) "Network design and transportation planning: Models and algorithms", *Transportation Science*, **18**, pp. 1-55.

Mason, L. G. (1983) "Network modernization with capital budget constraints", *Network Planning Symposium* (Networks '83), Brighton, U. K., pp. 5-9.

Mason, L. G. (1984) "An aggregate model for network modernization with constraints", *IEEE Transactions on Communications*, **COM-32**, No. 10, pp. 1073-1079.

Mason, L. G., and J.-P. Combot (1980) "Optimal modernization policies for telecommunications facilities", *IEEE Transactions on Communications*, **COM-28**, No. 3, pp. 317-324.

Mazzei, U., C. Mazzetti, and G. Roso (1988) "Economics of digital carriers and fiber optic systems in subscriber loop", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 133-137.

McGregor, P., and D. Shen (1977) "Network design: An algorithm for the access facility location problem", *IEEE Transactions on Communications*, **COM-25**, pp. 61-73.

Minoux, M. (1987) "Network synthesis and dynamic network optimization", *Annals of Discrete Mathematics*, **31**, pp. 283-324.

Mirzaian, A (1985) "Lagrangian relaxation for the star-star concentrator location problem", *Networks*, **15**, pp. 1-20.

Monma, C. L., and D. F. Shallcross (1986) "Methods for designing surviveable communication networks", Working paper, Bell Communications Research, Morristown, New Jersey.

Oda, T., Y. Wakahara, and M. Furuya (1986) "An efficient decision support system for facility planning of international transmission networks", *Proceedings of the Third International Networks Symposium* (Networks 86), pp. 37-42.

Pirkul, H. (1986) "Location of concentrators in designing local access networks", *Proceedings of IEEE INFOCOM '86*, pp. 148-154.

Rousset, G., and H. Cameron (1986) "Design and optimization of private data networks", *Proceedings of the Third International Network Planning Symposium*, pp. 50-54.

Sa, G. (1969) "Branch and bouhnd and approximate solutions to the capacitated plant location problem", *Operations Research*, 17, pp. 1005-1016.

Schneider, G. M., and M. N. Zastrow (1982) "An algorithm for the design of multilevel concentrator networks", *Computer Networks*, 6, pp. 1-11.

Sen, S., R. D. Doverspike, and M. S. Dunatunga (1989) "Unified Facilities Optimizer", Working paper, Systems and Industrial Engineering Department, University of Arizona, Tucson.

Sharma, R. L. (1983) "Design of an economical multidrop network topolocy with capacity constraints", *IEEE Transactions on Communications*, COM-31, pp. 590-591.

Shulman, A., and R. Vachani (1988) "A decomposition approach to solving the dynamic outside plant problem", Technical note, GTE Laboratories Incorporated, Waltham, Massachusetts.

Sinha, L. P., B. A. Freedman, N. K. Karmarkar, A. Putcha, and K. G. Ramakrishnan (1986) "Overseas network planning", *Proceedings of the Third International Network Symposium* (Networks '86), pp. 121-125.

Sirbu, M. A., and D. P. Reed (1988) "An optimal investment strategy model for fiber to the home", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 149-155.

Snelling, R. K., and K. W. Kaplan (1984) "Current and future fiber optics applications - Operating telephone company perspective", *Proceedings of IEEE Globecom '84*, pp. 604-607.

Spielberg, K. (1969) "Algorithms for the simple plant location problem with some side conditions", *Operations Research*, **17**, pp. 85-111.

Tang, D. T., L. S. Woo, and L. R. Bahl (1978) "Optimization of teleprocessing networks with concentrators and multiply connected terminals", *IEEE Transactions on Computers*, **C-27**, pp. 594-604.

Telephony (1985), July 8 issue, pp. 28-29.

The Economist (1987) "A survey of telecommunications: New lines for old", October 17, 1987, pp. 3-34.

Toth, A. G., E. Colombini, P. J. MacLaren, and R. K. Yates (1985) "Fiber in the local exchange network: A planning overview", *Proceedings of the IEEE International Communications Conference*, pp. 520-526.

White, P. E. (1986) "The broadband ISDN - The next generation telecommunications network", *Proceedings of the IEEE 1986 International Communications Conference*, pp. 385-390.

Woo, L. S., and D. T. Tang (1973) "Optimization of teleprocessing networks with concentrators", *Proceedings of the National Telecommunication Conference*, pp. 37C1-37C5.

Yaged, B. (1971) "Minimum cost routing for static networks", *Networks*, **1**, pp. 139-172.

Yaged, B. (1973) "Dynamic minimum cost network routing", *Networks*, **3**, pp. 193-224.

Yamamoto, Y., H. Yamamoto, and H. Oikawa (1984) "Access network design with optical fibers", *Proceedings of IEEE Globecom '84*, pp. 637-641.

Zadeh, N. (1973) "On building minimum cost communication networks", *Networks*, **3**, pp. 315-331.

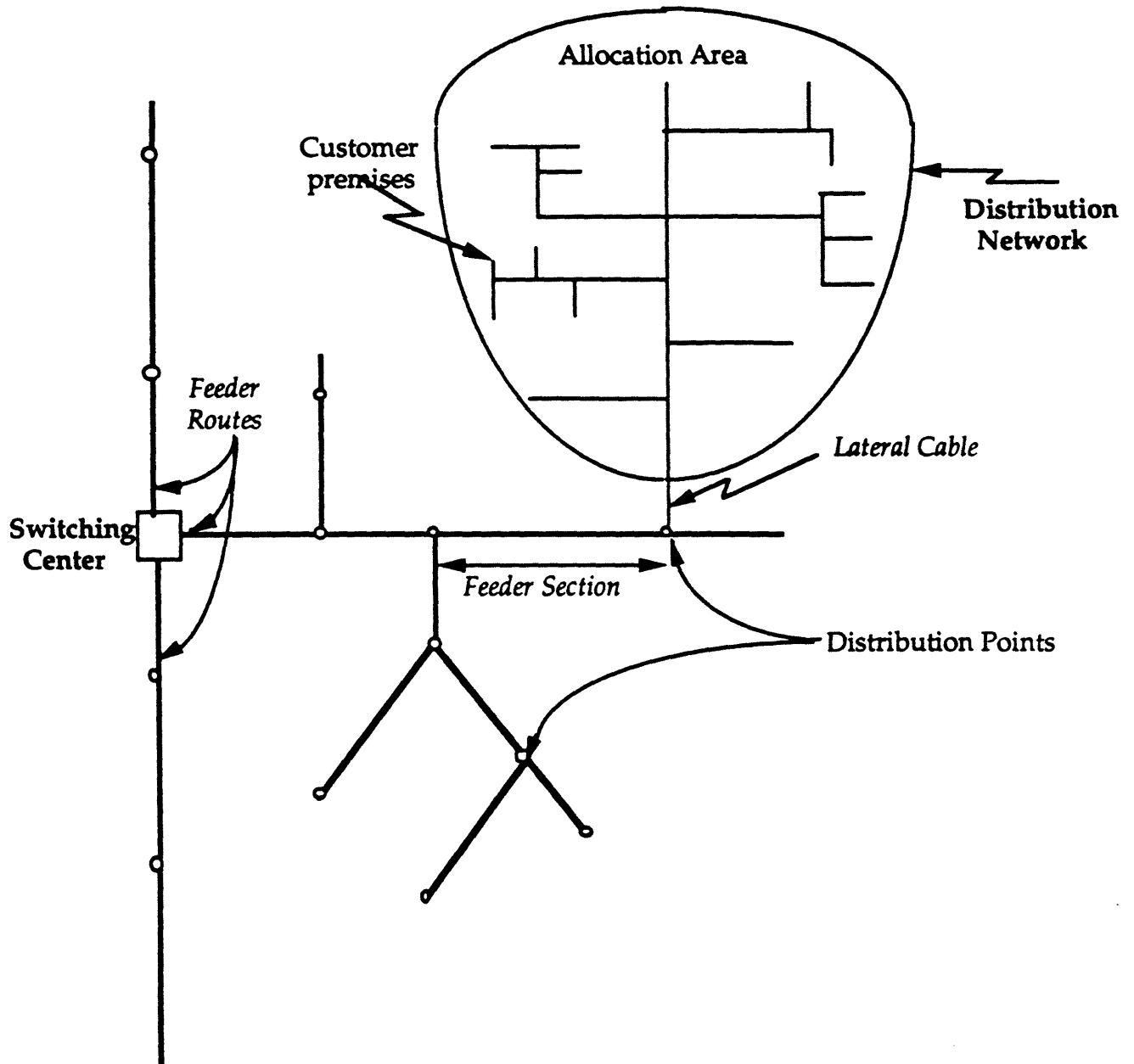Zadeh, N. (1974) "On building minimum cost communication networks over time", *Networks*, **4**, pp. 19-34.

Figure 1
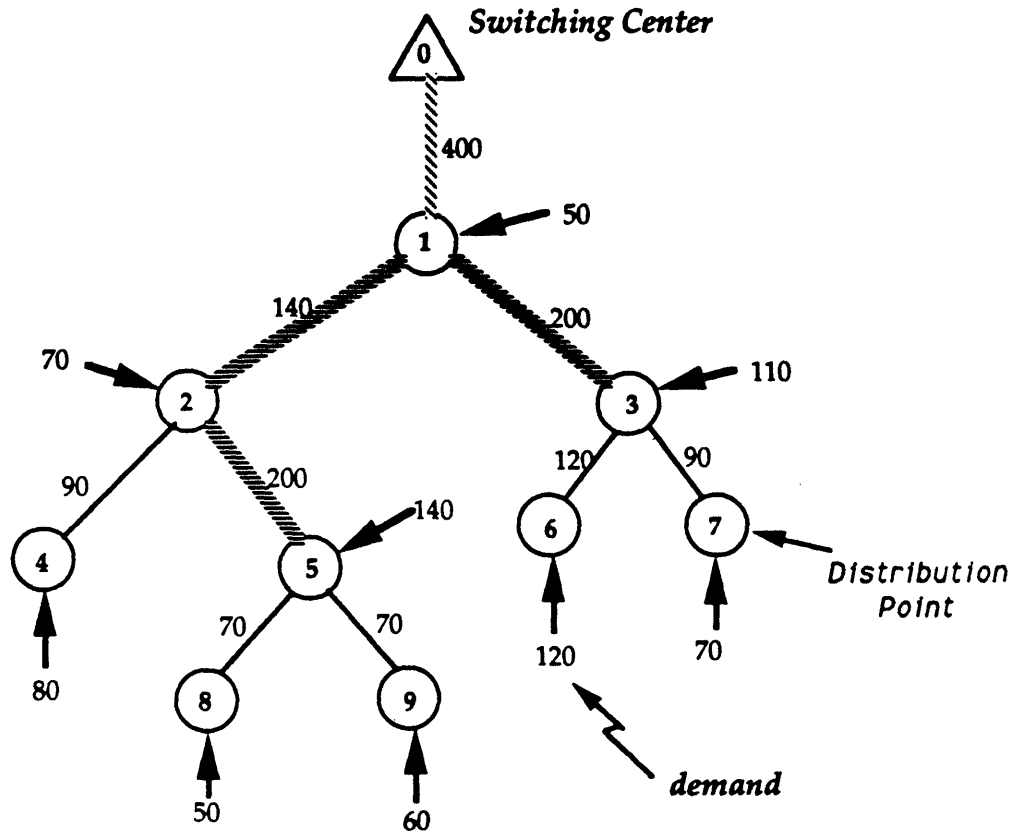## Hierarchical Structure of Telecommunication Networks

Gateway node

BACKBONE NETWORK

Switching Center

Distribution point

LOCAL ACCESS NETWORK

customers
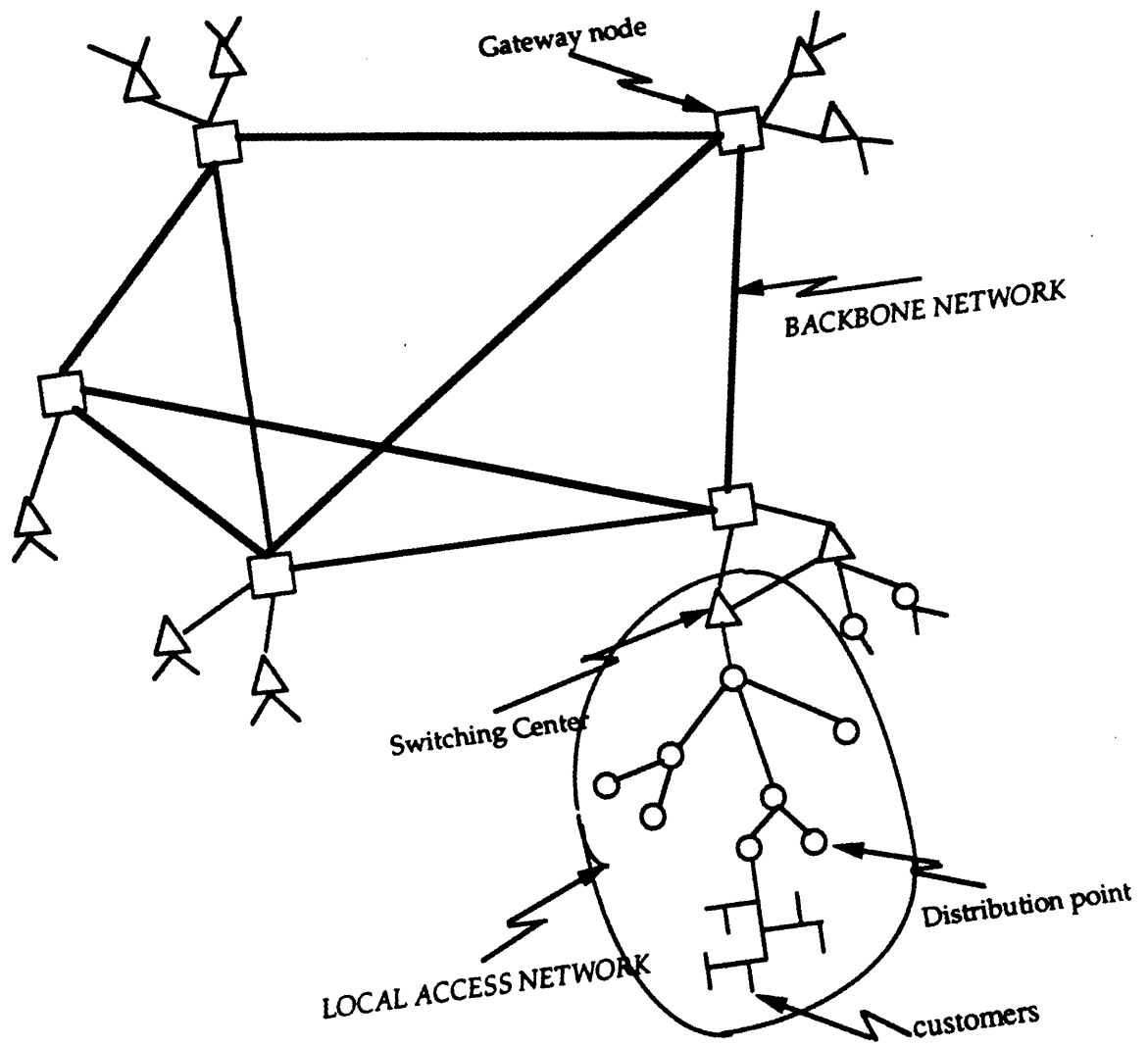
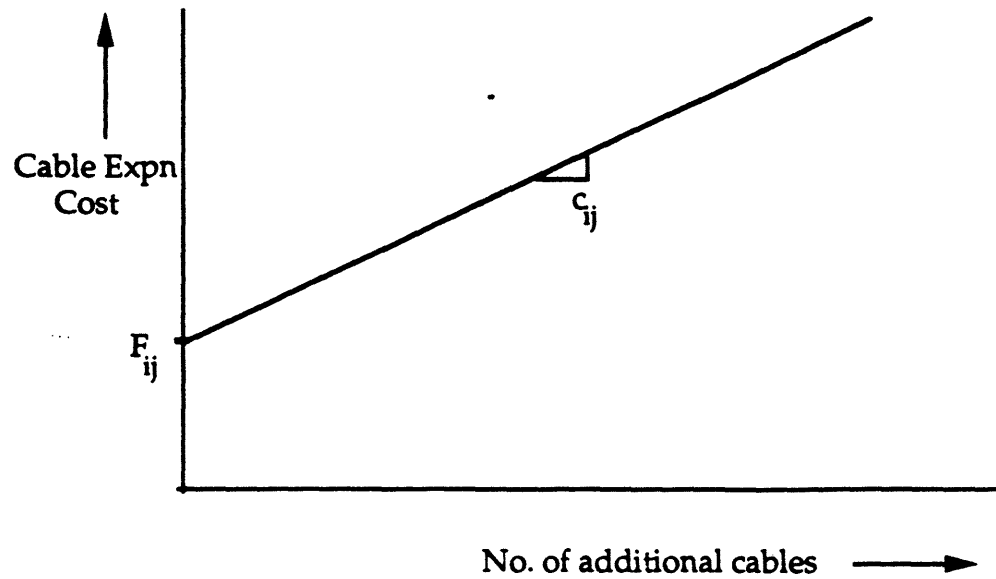# Figure 2

Typical Route of a Local Access Network

Figure 3A
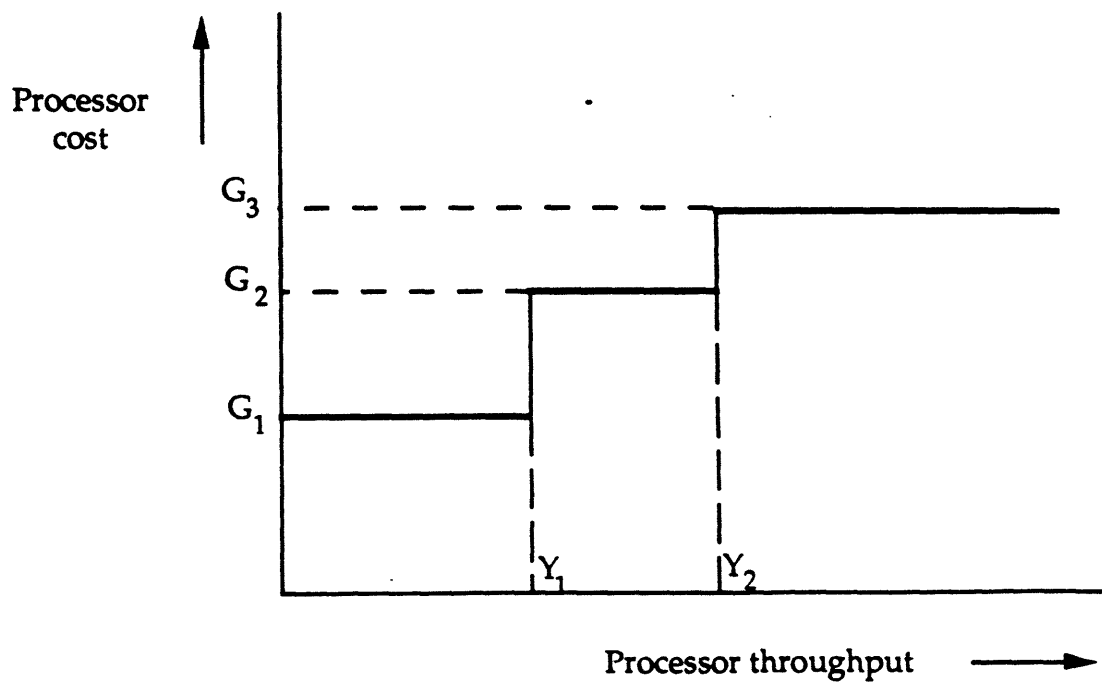
Local Access Network Planning Example

Numbers on edges denote current cable capacity
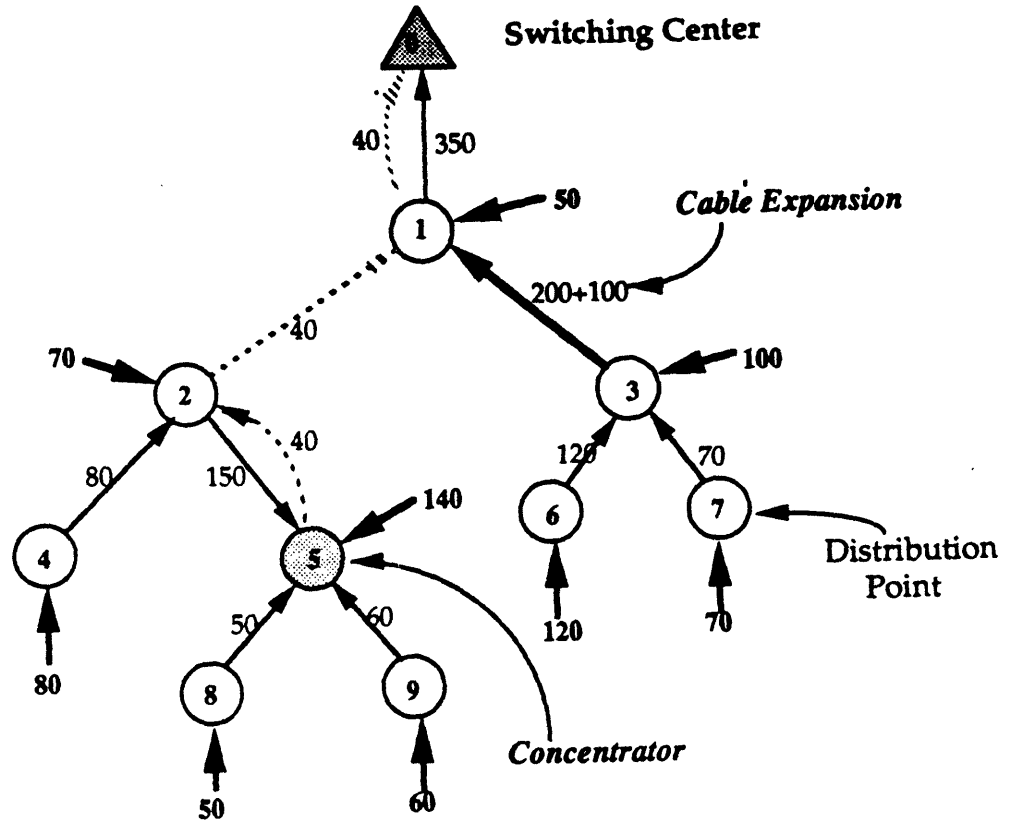
∿∿∿∿∿ BOTTLENECK edges
Cum. demand > Capacity

Figure 3b
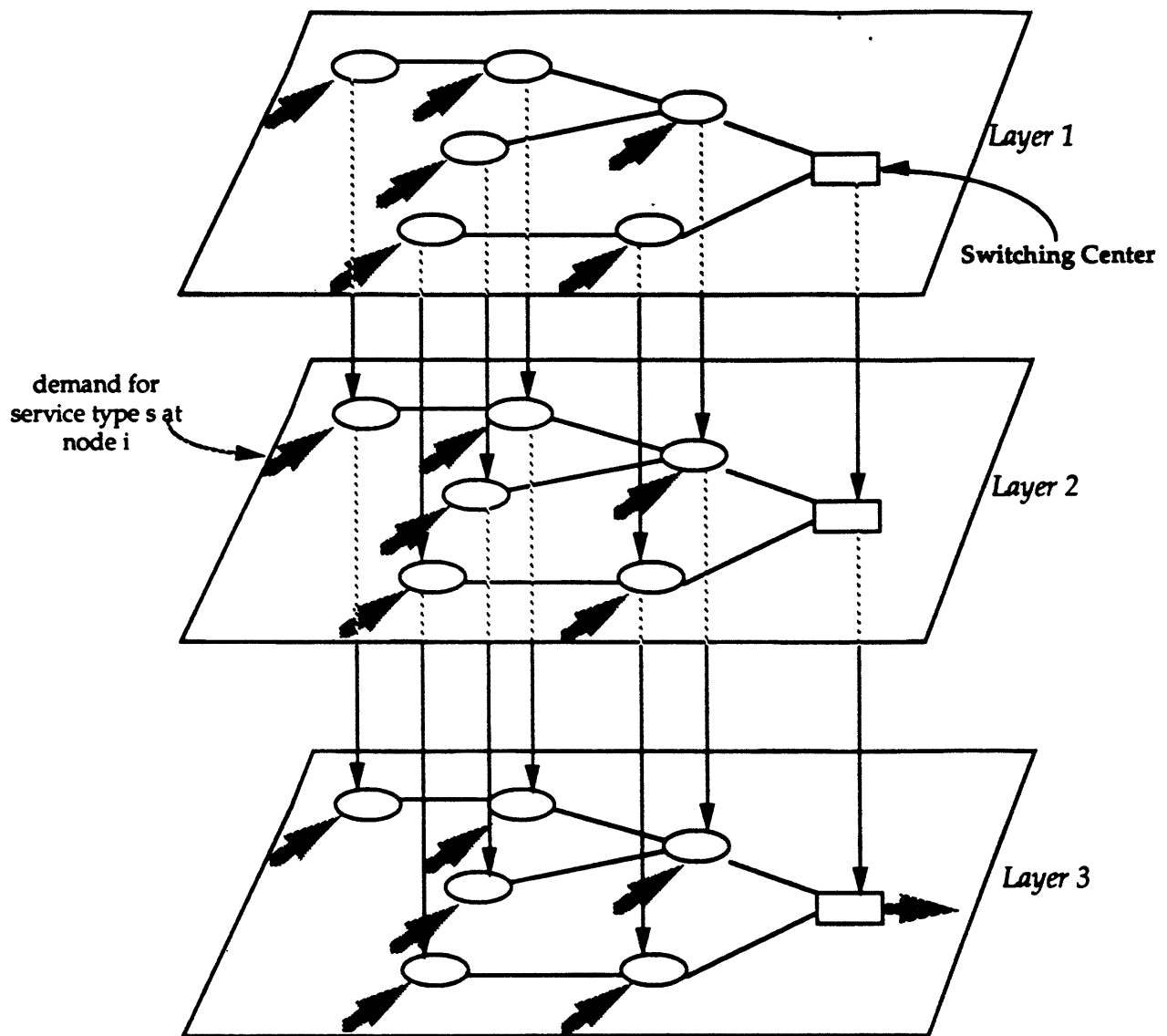### Cable Expansion Cost for Local Access Network Planning Example

Figure 3c
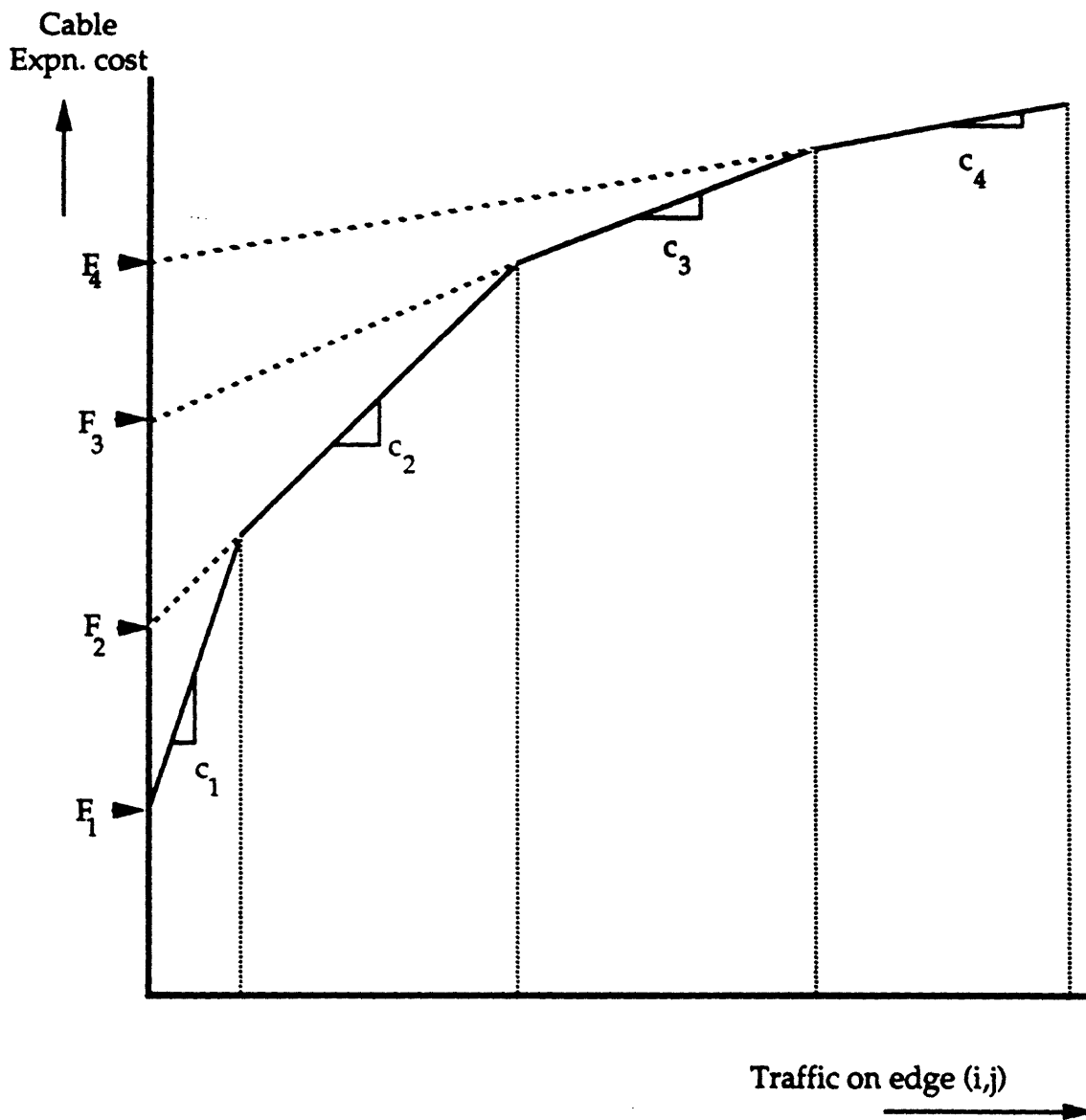### Processor Cost for Local Access Network Planning Example
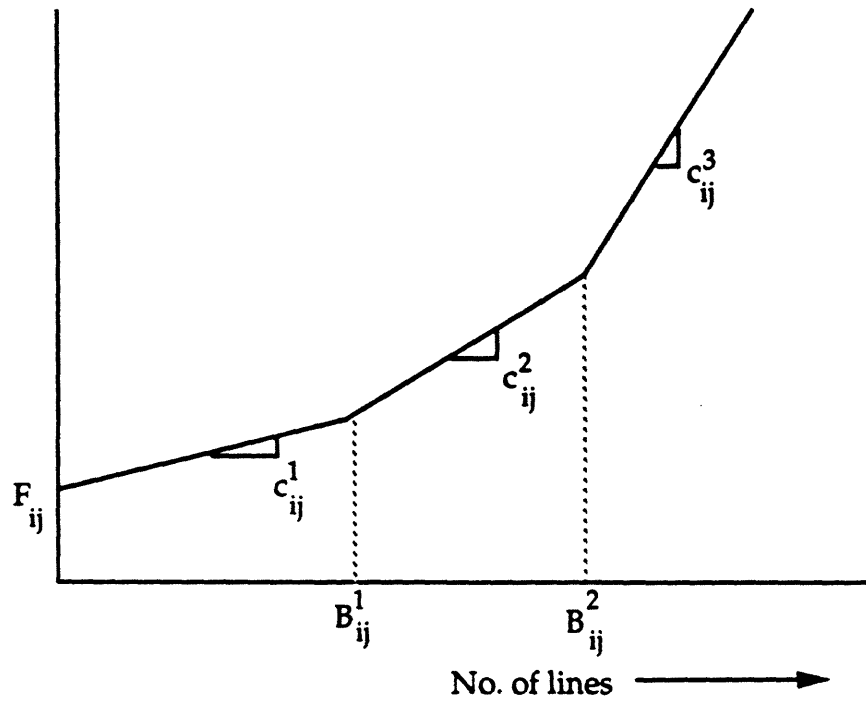
Figure 4
Sample Expansion Plan

.... Dotted lines show concentrated traffic

Numbers on edges denote number of lines used

Figure 5
Layered Network Representation

Layer 1

Switching Center

demand for
service type s at
node i

Layer 2

Layer 3

Figure 6
Cable Expansion Cost with Economies of Scale

Figure 7A
Convex Cable Expansion Cost

c_{ij}^3

c_{ij}^2

c_{ij}^1

F_{ij}

B_{ij}^1 B_{ij}^2

No. of lines ⟶

Figure 7B
Equivalent Network Representation

variable cost

Fixed cost $\longrightarrow$ $F_{ij}, c_{ij}^1, B_{ij}^1$ $\longleftarrow$ capacity

$0, c_{ij}^2, B_{ij}^2 - B_{ij}^1$

$0, c_{ij}^3, \infty$

i   j
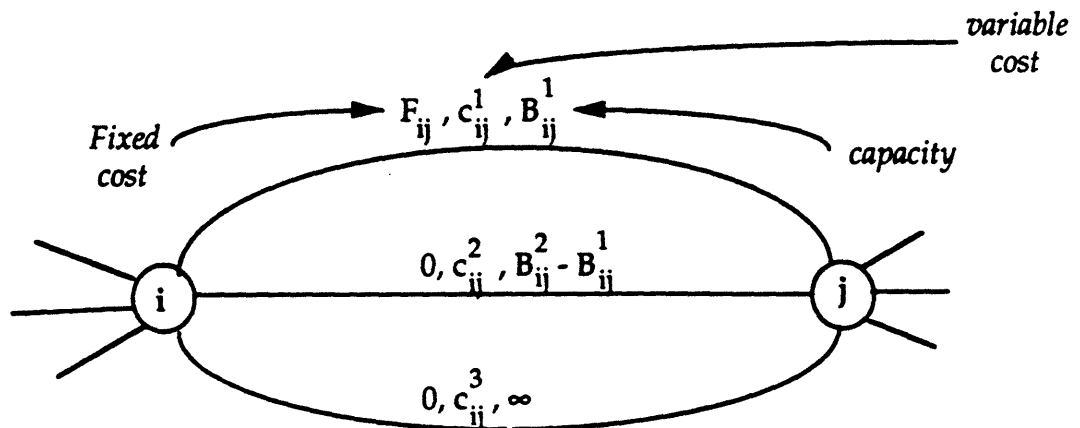
Figure 8A
General Cable Expansion Cost Function
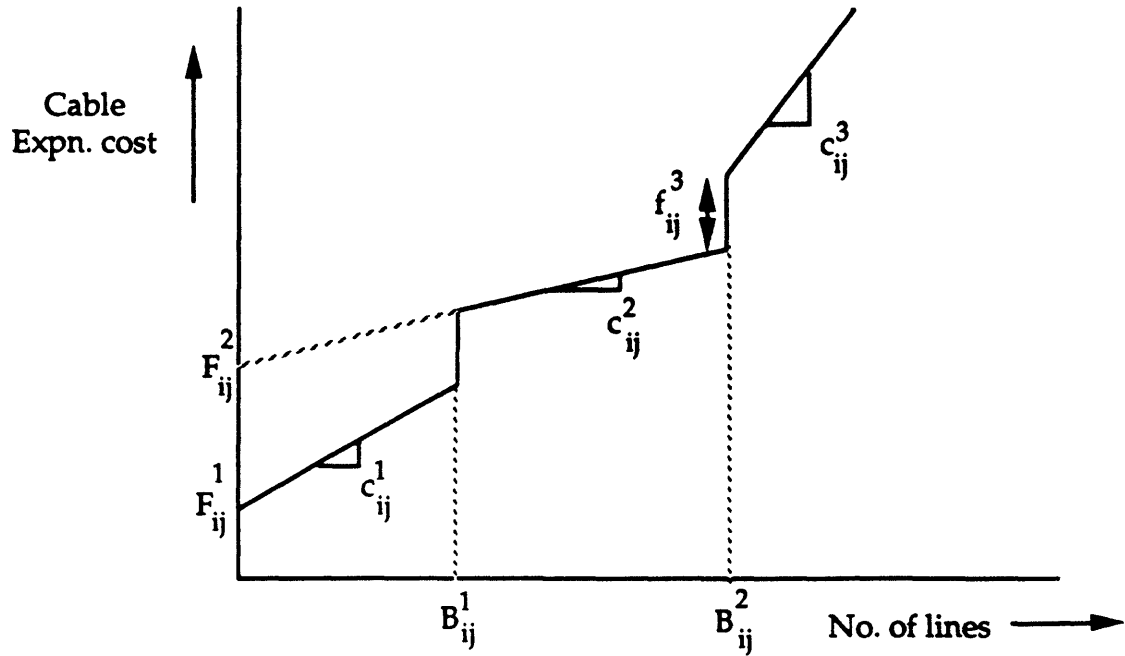
Figure 8B
Equivalent Network Representation

Figure 9
Line Network

Switching Center