Section IV

Computer and Communication Networks

Annals of Operations Research 33(1991)239–284

MODELS FOR PLANNING CAPACITY EXPANSION IN LOCAL ACCESS TELECOMMUNICATION NETWORKS*

A. BALAKRISHNAN☆ and T.L. MAGNANTI*
Sloan School of Management, M.I.T., Cambridge, MA 02139, USA

A. SHULMAN
GTE Laboratories, Incorporated, Waltham, MA 02254, USA

and

R.T. WONG
AT&T Bell Laboratories, Holmdel, NJ 07733, USA

Abstract

The rapid progress of communications technology has created new opportunities for modeling and optimizing the design 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 model and framework, and motivate different possible modeling assumptions; (iii) summarize various existing planning models in the context of this framework; and (iv) describe some new modeling approaches. The discussion in this paper is directed both to researchers interested in modeling local telecommunications systems and to planners interested in using such models. Our goal is to present relevant aspects of the engineering environment for local access telecommunication networks, and to discuss the relationship between engineering issues and 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 issues and discuss possible optimization approaches for treating them.

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

*This research was initiated through a grant from GTE Laboratories, Incorporated.
☆Supported in part by an AT&T research award.
*Supported in part by Grant No. ECS-8316224 from the Systems Theory and Operations Research Program of the National Science Foundation.

in communication facilities (estimated at around US$60 billion in 1980 in Bell System transmission facilities alone (AT&T Bell Laboratories [3]), and over US$100 billion in total for the US) 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 reviews and develops alternative modeling approaches for addressing contemporary design problems that arise in one major component of a telecommunication system: the local access network. The paper first sets a backdrop for our discussion by reviewing relevant technological developments as well as the evolution of the local access network.

In the next few years, the nature of services and the volume of demand in the telecommunications industry 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 (Coathup et al. [20], Dettmer [25], *Fortune* [32], Kostas [54], *The Economist* [82], Toth et al. [83]). Thus, ISDN combined with new switching and transmission technologies is expected to greatly stimulate network usage.

To accommodate the anticipated demand increase, telephone companies have initiated extensive modeling and planning efforts to expand and upgrade their switching and transmission facilities. 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 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 local access networks in the US employed digital switching (*The Economist* [82]). 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.

Because modernizing and expanding switching and transmission facilities requires enormous investments, telephone companies emphasize cost effectiveness in implementing selected expansion projects with high demand growth potential. For each project, 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 communication technologies has created additional decision alternatives and tradeoffs and, hence, new modeling challenges that did not arise in the traditional analog and copper environment. For instance, deploying concentrators and multiplexers in the local access network now provides an alternative means (instead of cable expansion) to increase network capacity. Consequently, network planners require new decision support models to identify cost effective expansion and modernization strategies.

This paper focuses on contemporary expansion planning models for the *local access* component (from the customer premises to the serving switching center) of public telephone networks. We do not address design issues, such as the blocking of potential transmissions or network vulnerability, that are more relevant for long-distance networks. Similarly, our models might not apply directly to data networks or rural networks which employ different technologies (for example, radio transmission and packet switching, respectively, in rural and data networks) and different criteria (for example, reducing packet delay in data networks).

This paper seeks 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 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.

The rest of this paper is organized as follows. Section 2 describes the evolution and engineering characteristics of local telecommunication 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 develops a general framework based on a layered network representation that encompasses a wide range of single-period local access network planning models, and motivates different possible modeling assumptions. Section 4 discusses several planning models in the context of 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 local access network, traces its evolution over the last few decades, and introduces some communications terminology. Our intent is to describe some important technological features so that we can represent them adequately in economic planning models.

2.1. THE COMMUNICATION NETWORK HIERARCHY

Most national telecommunications networks can be broadly divided into the three main levels shown in fig. 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 [56]);

(ii) the *inter-office* or *switching center* network within each city that interconnects *switching centers* (also called *local exchanges* or *central offices*) in different subdivisions (clusters of customers) and provides access to the gateway node(s); and

(iii) the *local access* network that connects individual subscribers belonging to a cluster to the corresponding switching center.

These three levels of the communication system hierarchy differ in several respects: the processing capabilities and amount of intelligence they contain, the technologies they employ, the services they perform, and their design criteria. For instance, the *local access network* typically has a tree configuration and contains a dedicated communication channel connecting each customer to the switching center. Most local access networks (approximately 80% in the US) currently use analog transmission on copper cables and do not contain electronic devices. In contrast, the *long-distance network* has a relatively dense topology, providing multiple communication paths between gateway nodes which contain intelligent hardware to perform switching, traffic compression (concentration), and some service functions (such as directory assistance). The long-distance networks in the US are almost completely digitized and employ high-frequency transmission using fiber optics, microwave, and satellite communications. The *inter-office network* links 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 inter-office network contains limited intelligence for routing incoming messages to the appropriate downstream switching centers or 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 different levels are interdependent. For instance, the number of customers and volume/mix of traffic assigned to each switching center determines the desired switching and transmission capacities on the inter-office network. However, because of differences

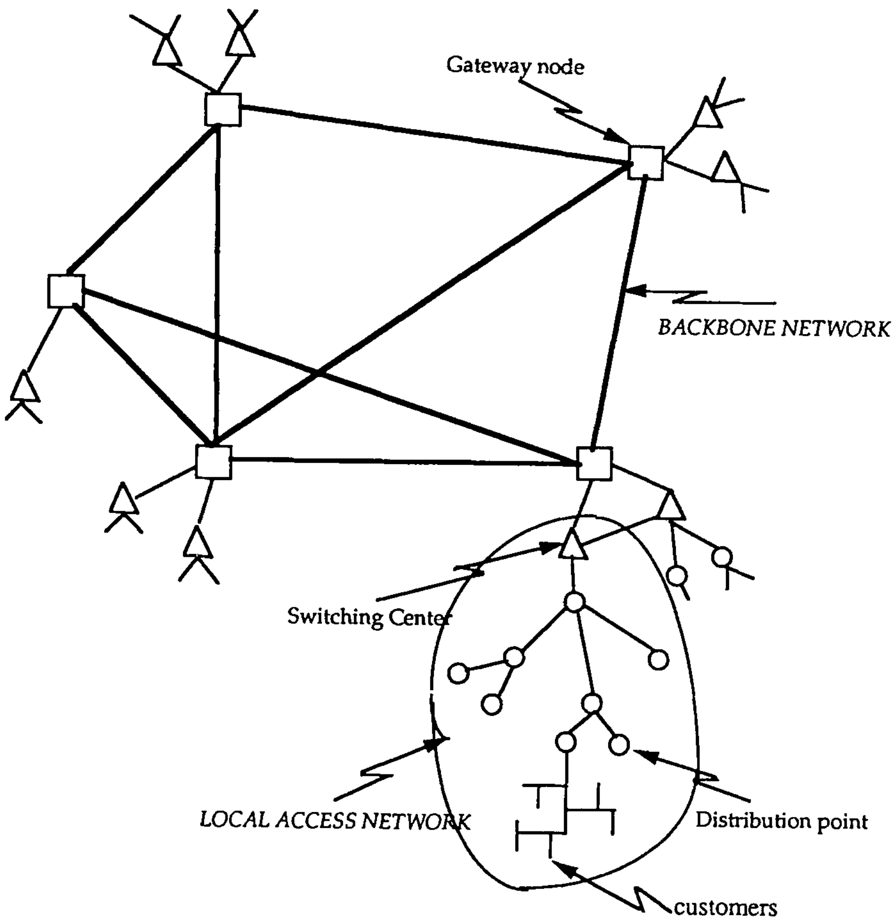


Fig. 1. Hierarchical structure of telecommunication networks.

in ownership and to simplify the planning task, analysts decompose the overall planning problem by considering each level separately (see, for example, Dawson, Murphy and Wolman [24]). In this paper, we focus on decision models for designing and expanding the local access component.

2.2. EVOLUTION OF THE LOCAL ACCESS NETWORK

The local access network (also called the *outside plant, local loop,* or *local exchange network*) links individual customers to a switching center that interconnects them for intra-exchange communications, and also serves as an interface to higher levels in the network hierarchy. Like the overall communication system, this network is also hierarchical; it consists of three levels referred to as *routes, feeder networks,* and *distribution networks.*

A *route* is a portion of the local access network containing all customer nodes that communicate with the switching center via a common link incident to the center. Figure 2 shows the structure of a route in a local access network. Each switching

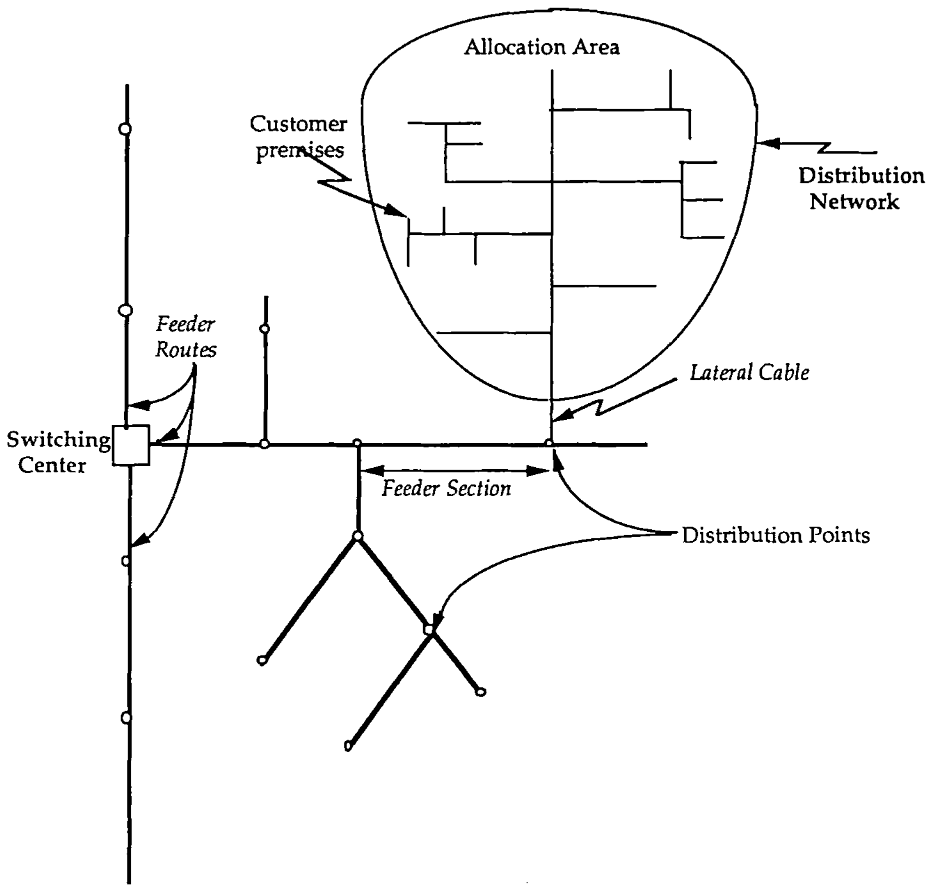


Fig. 2. Typical route of a local access network.

office might serve as the termination point for 3 to 5 routes (Koontz [52]); the Bell System contains around 40,000 such routes (Ciesielka and Douglas [18]). Typically, analysts consider each route independently for capacity planning purposes.

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 of varying gauges that are either buried, installed in ducts, or mounted on poles, and are accessible at intermediate points. The number of distribution points assigned to a switching center varies from 20 to 200. The segment of cables between two adjacent distribution points along the route is often called a *feeder section.* The feeder network has a tapered 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, sometimes called an *allocation area* (Gibson and Luber [38]), typically has a diameter of a few thousand feet, and may include 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 [39] for a more comprehensive description of local telecommunication networks.)

Traditionally, distribution networks are designed for ultimate demand (which is relatively small) in order to exploit economies of scale and to avoid subsequent disruption of service for laying new cables. On the other hand, feeder networks are designed to meet only medium-term demand; telephone companies periodically review and increase feeder capacity to accommodate demand growth and customer movement (Ciesielka and Long [19], Elken [28], Freidenfelds and McLaughlin [34]). This paper focuses on 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 (see also Dawson et al. [24]).

Stage 1: The basic feeder network

The basic feeder network employs analog transmission at the voice frequency of 4 kHz over copper cables (twisted wire pairs). It uses a dedicated line to connect each customer to the switching center. Physically, the line for a customer might consist of wire segments (possibly with different gauges) in each downstream feeder section that are joined at the intermediate distribution points.

In this setting, a principal design concern 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, sometimes called *Resistance Design* (Ciesielka and Douglas [18]), consists of selecting a cost-effective combination of wire gauges for each feeder section to satisfy all maximum resistance requirements.

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.* Any section where demand exceeds the available cable capacity is said to have *exhaust.* The feeder planning exercise considers two strategies to relieve exhaust when customers move or the number of customers increases: (i) feeder cable reallocation, and (ii) feeder cable expansion.

Given the projected changes in the 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 upstream distribution points in order to delay cable expansion. Gibson and Luber [38] describe a heuristic feeder allocation method; Elken [28] formulates the reallocation task as a separable convex programming problem, and proposes an iterative procedure that solves a sequence of linear programs.

Feeder expansion models (e.g., Bulcha et al. [13], Freidenfelds and McLaughlin [34], Koontz [52]) determine the number of additional cables to install over time in order to relieve the projected exhaust at minimum total discounted cost. For tree networks with cable expansion as the only available method for relieving exhaust, each feeder section can be analyzed independently for capacity planning purposes (since the number of customers served by each distribution point uniquely determines the cable requirements in every section). Thus, physical pair facility relief models used in this context do not incorporate any spatial coupling between sections. Freidenfelds and McLaughlin [34] formulate a multiperiod capacity expansion model to find the optimal mix of cable gauges in each section; they describe a heuristic solution method for this model.

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* or *remote electronic* devices, i.e., *multiplexers, concentrators* and *remote switches*, 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 in frequency division multiplexing, and 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, combining 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 (and so signals might be blocked): rather, the output channels are dynamically assigned to input lines as the need arises. The ratio of incoming to outgoing *channels* is called the *concentration ratio. Remote switches* are decentralized, smaller vesions of the main switching center; they perform local switching functions to interconnect all customers who communicate through them, and also compress the traffic destined to the main switching center.

In local access network applications, remote electronic devices enable multiple users to share the same physical line on the feeder network, thus providing an alternative method to relieve exhaust as demand increases. They also eliminate circuit resistance restrictions, and permit the use of fewer and less expensive wire gauges. Multiplexers, concentrators, and remote switches 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 and remote switching reduce the number of cables required in downstream feeder sections, these cables must now handle the higher output frequencies of the electronic devices. Conventional copper cables (twisted pairs) have a limited effective bandwidth (around 150 kHz). Higher frequency signals (150 kHz to 2 MHz) require either coaxial cables or *conditioned* (or *groomed*) copper cables, i.e., twisted wire pairs with intermediate repeaters (which are electronic devices to eliminate signal distortion); very high frequency signals (over 2 MHz) require fiber optic cables. Often, existing ducts (built for copper cables) can accommodate these enhanced transmission media as well.

For the second generation local access network with electronic devices, the planner must consider various choices for locating, sizing, and timing the installation of remote electronics (multiplexers, concentrators, switches) as well as the conventional option of physical pair facility relief (i.e., increasing cable capacities in different feeder sections). Furthermore, unlike the older technologies, the new traffic compression devices introduce spatial couplings, i.e., we can no longer consider each feeder section in isolation since higher demand at a distribution point does not necessarily translate into increased cable capacity requirements on every downstream feeder section.

Stage 3: Fiber in the local access network

Many telephone companies are currently planning to introduce fiber optic technology in local access networks because of its extremely high bandwidth. Fiber optic or lightwave transmission facilities consist of a fiber cable connecting a pair of *fiber optic terminals* (or *fiber terminating equipment*) that convert electrical (analog or digital) signals into very high frequency optical signals, and also perform optical coupling and multiplexing functions. The bandwidth of fiber cables (some with a transmission capability of over 1 Tera bits per second) is effectively unlimited for local network applications; therefore, they can accommodate a large number of multiplexed high-frequency channels. Indeed, the electronic circuitry in the fiber terminating equipment is currently the main factor limiting the number of channels that can be multiplexed on fiber. For local access networks, 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.

Except for a few experimental networks, telephone companies have not deployed fiber optic transmission extensively in the local access network. The characteristics, capabilities and deployment of this technology, and even the network configuration plans, are constantly changing (see, for instance, Anderson [1], Carse [15], Ensdorf, Keller and Kowal [29], Mazzei, Mazzetti and Roso [64], Snelling and Kaplan [80], Toth et al. [83]). Researchers have proposed several competing topologies, such as the ring, loop, and mesh architectures, for fiber-based local access networks (e.g. Campbell [14], Garbanati and Palladino [35], Sirbu and Reed [79], White [85], Yamamoto, Yamamoto and Oikawa [87]). These topologies seek to reduce network vulnerability, an issue that is becoming increasingly important for fiber networks since damage to even a single cable can affect service to a large number of customers.

This paper assumes that, for topological design 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 deploying it, network planners might require more sophisticated models to distinguish fiber optic transmission from conventional electrical transmission.

Table 1 summarizes the technologies and expansion options we have discussed in this section. The next section formalizes the feeder network planning problem, and motivates several possible modeling assumptions; we will subsequently use these assumptions to differentiate various modeling approaches.

Table 1

Local access network expansion options

Technology stages	Technology	How technology provides additional capacity
1. Basic feeder network	Copper cables	Physical pair relief • Reallocate cables (feeder reallocation) • Expand cables (feeder expansion)
2. Feeder network with remote electronics	• Multiplexers	• Compresses or interleaves signals
	• Concentrators	• Compresses signals with dynamic channel assignment
	• Remote switches	• Provides local switching
3. Fiber in local access network	Fiber cables Fiber optic terminals	Converts electrical signals to very high frequency optical signals (with multiplexing)

3. Local access network planning: Problem definition, modeling framework and assumptions

This section presents a general conceptual and modeling framework for studying local access (feeder) network planning, defines the scope of models that we will subsequently consider, and identifies some key assumptions that distinguish different modeling approaches for this problem. We present an example to clarify the decisions and tradeoffs in the planning task, and to illustrate the modeling implications of traffic compression and cable expansion options. We then develop a general layered network representation that captures the important issues and tradeoffs in local access network planning.

The feeder capacity planning exercise begins with a forecast of telecommunication demand (based on new construction and customer movements) at each distribution point for the duration of the planning horizon. 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 US 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, for some types of data transmission) of the basic DS0 rate; for example, the DS1, DS2, and DS3 rates are, respectively, 24, 96, and 672 times the DS0 transmission rate.

The *objective* of the planning exercise might be to either satisfy *all* projected demand at *minimum* total (investment plus operating) discounted *cost*, or *selectively* satisfy demand to *maximize total profit*. In this paper, we focus on the more common cost minimization perspective used by telephone companies. Furthermore, we restrict our discussions to *single-period* or *static* models rather than *multi-period* models. Multi-period models can account for temporal couplings caused by economies of scale (e.g., installing excess capacity in anticipation of future demand increases); however, they are much more difficult to solve than static models that only satisfy demand in the terminal year of the planning horizon. Studying static models might possibly give us insights about the more general problem. For instance, single-period solution algorithms could serve as building blocks for multi-period versions (see, for example, Shulman and Vachani [78]). Or, as Minoux [66] 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.

To *meet the demand* for different services, the network design must provide adequate and appropriate traffic processing (multiplexing, concentration, and switching) and transmission facilities from each distribution point to the switching center. In general, different services and processing devices require different transmission rates and, hence, different transmission media (twisted wire pairs, groomed copper cables, coaxial cables, and fiber optic cables). For example, video signals cannot be transmitted over twisted pair copper cables; similarly, remote electronic devices with high multiplexing ratios require enhanced media to carry the high frequency output signals. The local access network design must account for these bandwidth specifications and provide compatible communication resources.

In addition to providing adequate capacity to meet the projected demand, the local access network configuration must also satisfy various *technological and policy restrictions*. For example, telephone companies might wish to provide multiple paths to some preferred large-volume business customers. Similarly, to ensure adequate transmission quality (particularly for data transmission applications) and to facilitate future expansion and modernization, designers might specify a maximum permissible distance (some companies use a limit of 12 kilofeet) between each customer and the nearest electronic device; we refer to this type of constraint as a *proximity restriction*.

Thus, the single-period local access network planning problem has the following ingredients:

Given (i) the projected (terminal year) demand for different services at each distribution point,
(ii) the current processing (switching) and transmission capacities at each location, and
(iii) the costs of installing, expanding and operating new processing and transmission facilities,

find the cost minimizing expansion plan that meets the demand and satisfies all technological and policy restrictions.

The optimal expansion plan should specify (a) the location and size of various network enhancements (i.e., addition or expansion of transmission media and remote electronic devices), and (b) the routing of traffic from each distribution point to the switching center.

Observe that we ignore special *information* processing steps (such as database queries) required by some services. Currently, all information processing occurs at higher levels of the communication hierarchy (i.e., in the inter-office or backbone network); the local access networks perform only *traffic* compression. Future designs might even decentralize certain information processing functions to the local access network. Also, our single-period model does not consider tradeoffs between overlay and replacement strategies for new technologies. In a multi-period setting, the planner must also decide whether to introduce new digital technology by initially overlaying existing analog technology, or by immediately replacing the analog components. Combot and Epstein [21], Combot, Tsui and Weihmayer [22], Hoang and Lau [45], Kopp [53], Mason [61, 62], and Mason and Combot [63] consider these modernization issues for inter-office network planning. Before presenting a general modeling framework for local access network planning, we discuss a simple example that illustrates the basic decisions and tradeoffs that the network planner must address.

3.1. EXAMPLE

Figure 3(a) illustrates a local access network problem with a single service type. The given network has a tree structure, consisting of a single medium (say, copper cables) with existing capacities as shown in the figure. However, this capacity is insufficient for the projected demand level. The heavy shaded lines represent feeder sections with projected exhaust (i.e., capacity shortfall). The projected exhaust represents the required amount of cable capacity expansion in a conventional physical

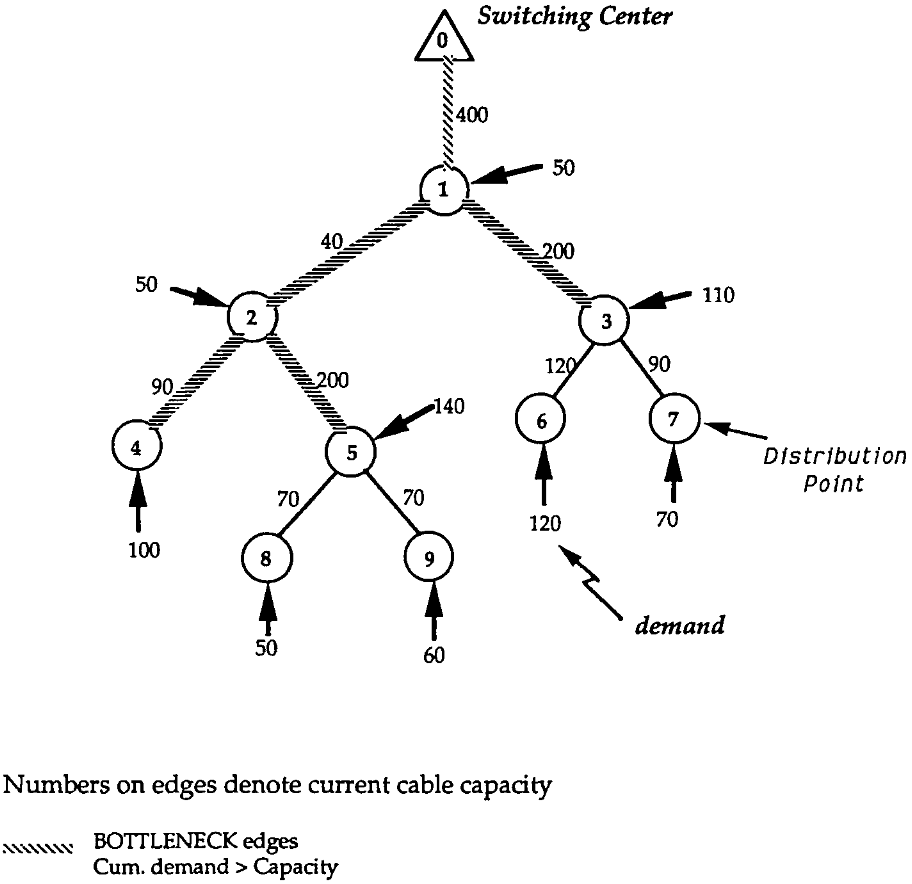


Fig. 3(a). Local access network planning example.

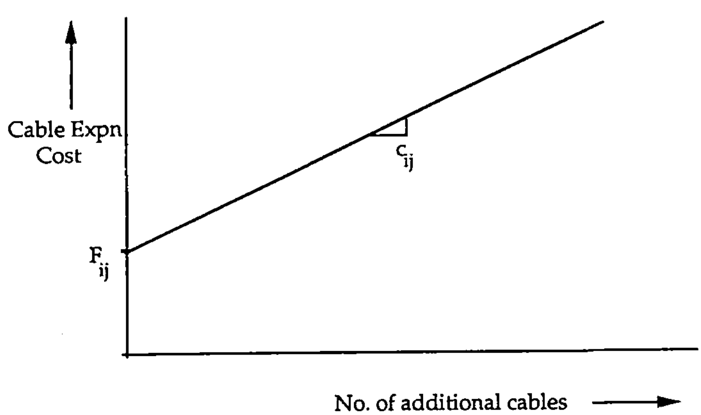


Fig. 3(b). Cable expansion cost function.

pair relief strategy (that does not employ traffic processing options). Figure 3(b) shows a representative cost function, consisting of a fixed charge plus a (constant) per unit cost, for expanding cable capacity along the feeder section between distribution points i and j.

The network in this example does not have any existing processors. Suppose that the planner has one available processor type with a 10:1 traffic conversion ratio, i.e., the processor compresses the traffic entering on every set of 10 incoming lines onto 1 outgoing line. We assume for simplicity that the existing copper cables can transmit both the base rate signal at which the service originates and the processor's compressed output signal. Figure 3(c) shows an illustrative processor cost function.

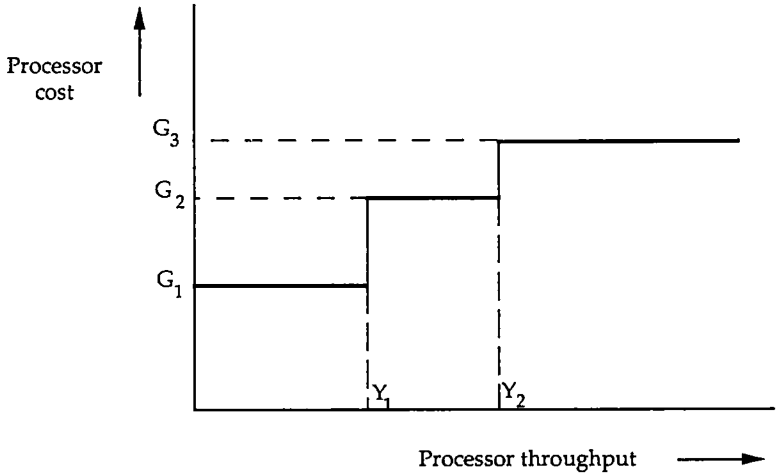
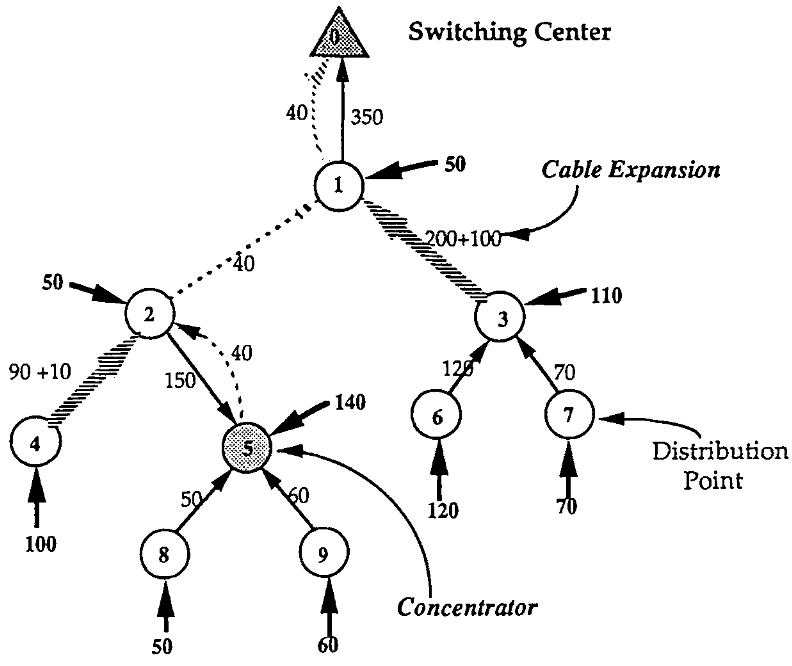


Fig. 3(c) Processor cost function.

This cost function might correspond to three processor types with differing costs (G_1, G_2, and G_3) and capacities (Y_1, Y_2, and ∞). The planner must select processor locations and capacities, and decide what cable section to expand by how much, in order to satisfy the demand at minimum total cost.

Figure 4 shows one possible expansion plan for the network example of fig. 3. This plan entails installing a processor with a (input) capacity of 400 units at node 5, and expanding the cable segments along sections (3,1) and (4,2) by 100 and 10 units, respectively. The processor at node 5 compresses 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. All the other nodes transmit signals at the base (unconcentrated) rate to the switching center. By installing the processor at node 5, we have relieved



⋯⋯ Dotted lines show concentrated traffic

Numbers on edges denote number of lines used

Fig. 4. Sample network expansion plan.

the projected exhaust on edges (2,5), (2,1), and (1,0) (the physical pair relief strategy would have expanded cable capacities in these feeder sections). Observe that we permit traffic flow in either 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, and 40 units of concentrated traffic flow in the opposite direction from 5 to 2 (to the switching center); thus, the 200 available lines in section (2,5) can accommodate both these flows. Also note that the expansion plan shown in fig. 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 backfeed.

The example of figs. 3 and 4 illustrates two tradeoffs in local access network planning:

(i) *Processor installation versus 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). This strategy is more economical if the 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).

(ii) *Centralized versus distributed processor location*: For example, should we locate two small processors, one at node 4 (with a capacity of 150 units, serving nodes 2 and 4) and the other at node 5 (with a capacity of 250 units, serving nodes 5, 8, and 9), instead of a single large processor at node 5? The two-processor solution avoids cable expansion on section (4,2); however, due to economies of scale, its total processor cost is likely to be higher. In general, the total cable expansion cost might possibly increase as the number (and size) of processors decreases, while installing fewer, but larger, processors reduces the total processing cost. The planning model must address this tradeoff between exploiting economies of scale in processor costs and avoiding transmission capacity expansion through a decentralized processor location strategy.

For illustrative purposes, we considered a simplified example with a single service, a single processor type, and a single medium that can transport both compressed and original traffic. Additional complexities arise when we consider multiple processor types, multiple processing steps in series, and multiple transmission media. In the next two sections, we develop a layered network representation for this more general problem. This representation serves as a framework for comparing various modeling approaches for local access network expansion planning. It encompasses a wide range of existing and proposed transmission and processing technologies, cost structures, and network topologies.

3.2. MODELING PRINCIPLES

This section discusses the modeling elements for representing local access network planning problems with multiple transmission rates, service types, processor types, and transmission media. In section 3.3, we develop a conceptual framework based on a multi-layer network representation that captures the interrelationship between the three main modeling elements: (i) *customer demands* for different services that must be satisfied; (ii) *transmission facilities* for carrying signals; and (iii) *traffic processors* (e.g., multiplexers, concentrators, remote switches or fiber optic devices) that can compress signals. As mentioned previously, the network expansion plan must match the bandwidth (i.e., transmission rate or frequency) specifications of these three elements. The layered network representation ensures this compatibility by separately identifying the traffic flows at each transmission rate. We first discuss how the transmission rate serves as a common link between customer demands, transmission facilities, and traffic processors.

We assume that the local access network employs a discrete set of "standard" transmission rates. For instance, telephone companies in the US use four standard *digital* transmission rates (expressed as multiples of *bits per second* or *bps*) labeled DS0, DS1, DS2, and DS3; these rates correspond, respectively, to 64 Kbps, 1.536 Mbps, 6.144 Mbps, and 43.008 Mbps. (A new set of standard (SONET) high frequency transmission rates (over 50 Mbps), denoted as OC1, OC2, etc., is emerging for broadband architectures.)

3.2.1. Customer demands

Customer demands are forecasted by service type (voice, data, video) at each distribution point. Each service type originates at a *basic rate*. For instance, certain types of video services require a basic rate of 1.5 Mbps (the DS1 rate). This rate represents the minimum frequency at which the service must be transmitted; the signals may be multiplexed to higher frequencies, if necessary. The terminal year demand for each type of service is expressed as the number of required channels at the basic rate. The final design should provide the required number of basic rate channels (or an equivalent number of higher rate channels) for each service type from every node to the switching center. Observe that, since we do not consider any unique information processing requirements for different service types, we can effectively aggregate all services requiring the same basic rate into a single service type.

3.2.2. Transmission facilities

We differentiate transmission facilities according to the medium or cable type (such as twisted wire pairs, groomed copper cables, coaxial cables, and fiber optic cables). Each transmission medium can handle a limited set of transmission rates. For example, copper cables can transmit DS0 and DS1 signals, while the DS2 and DS3 rates require coaxial or fiber optic cables. Some transmission media (e.g., copper cables) are installed in sections; to connect two non-adjacent nodes, the cable sections must be joined at intermediate distribution points. We refer to these media as *sectional* media. In contrast, *point-to-point* media (e.g., fiber cables) provide direct connections without intermediate connectors; however, they may use the same physical infrastructure, namely, the underground ducts or trenches in each intermediate feeder section. Most high frequency traffic requires point-to-point cables.

Transmission capacities for each medium are specified by the number of lines of that medium. In general, the total cost to install or expand the transmission capacity between any pair of nodes has two components: a *separable* cost component pertaining to each *individual* medium, and a *joint* or shared cost that is common to several media. Joint costs arise because different media share the same infrastructure (e.g., underground ducts). Because of scale economies, both the separable and joint cost functions are likely to be concave; for example, they might consist of a fixed expense and a cost per unit of traffic.

3.2.3. Traffic processors

As explained previously, traffic processors combine several incoming (lower frequency) signals into a single outgoing (higher rate) signal. The economic models that we consider characterize each processor type by three parameters: its *input rate* (i.e., frequency of input signals), *output rate*, and *conversion ratio.* 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 traffic *compression* ratio, i.e., the ratio of output *rate* (or frequency) to input *rate*, because of provisions for spare outgoing lines and differences between the number of input and output channels (as noted previously, concentrators have fewer output channels than input channels since they employ dynamic channel allocation). Effectively, the input and output rates determine the type of input and output transmission media that the processor requires, while the conversion ratio determines the number of physical lines of the input medium per outgoing line.

As a convention, we specify processor capacities in terms of the number of input lines. Installing new processors or expanding existing capacities entails *processor costs* which might vary by location, processor type, and the required additional capacity. For instance, processor costs might include fixed expenses (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.

A specific commercially available traffic processing device, namely, the SLC-96 system (Ciesielka and Douglas [18]) illustrates these concepts. The SLC-96 is a modular digital carrier/concentrator system 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. The input rate of 4 kHz is equivalent to a DS0 digital 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 so-called 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 * 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 × 2pairs/line}) = 16. One version of the SLC-96 system permits stacking of up to ten modules (to provide service for up to 960 customers).

3.2.4. An illustrative system

Table 2 summarizes the relationships between transmission rates, customer demands for different service types, transmission facilities, and traffic processors. For ease of illustration, we consider only three transmission rates – DS0, DS1, and DS2 – and assume that each rate has a unique corresponding medium (twisted wire

Table 2

Illustrative system

Layer	Transmission rate (frequency)	Service type	Transmission medium	Conversion ratios* (Compression ratios☆) of traffic processors		
				Type 1	Type 2	Type 3
1	DS0 (64 Kbps)	Voice	Twisted pair	12:1 (24:1)		48:1 (96:1)
2	DS1 (1.536 Mbps)	Data	Coaxial		2:1 (4:1)	
3	DS2 (6.144 Mbps)	Video	Fiber			

*Conversion ratio = number of *input lines*/number of *output lines*.
☆Compression ratio = *output frequency*/*input frequency*.

pair, coaxial cable, and fiber cable, respectively). The example considers three service types – voice communication, high-speed data, and video service – whose basic rates are, respectively, DS0, DS1, and DS2. In this example, the designer can use three types of processors: a type 1 processor to compress DS0 signals to the DS1 rate (i.e., with a DS0 input rate and a DS1 output rate) with a conversion ratio of 12; a type 2 processor to compress DS1 signals to the DS2 rate with a conversion ratio of 2; and a type 3 processor with a conversion ratio of 48 to directly compress DS0 traffic to the DS2 rate. Observe that the conversion ratios differ from the traffic compression ratio; for instance, DS1 signals represent a 24-fold *compression* of the DS0 rate, while the type 1 processor has a *conversion* ratio of only 12 (i.e., it requires 1 outgoing line for every 12 incoming lines). Also, compressing DS0 signals directly using a type 3 processor is more effective than using a type 1 and a type 2 processor in tandem; for example, with 480 incoming DS0 lines, the type 3 processor requires only 480/48 = 10 outgoing DS2 lines, whereas the type 1 – type 2 combination requires 480/(12 * 2) = 20 DS2 lines. The next section presents a layered network representation to model the interrelationships shown in table 2.

3.3. LAYERED NETWORK REPRESENTATION

Let $G:(N, E)$ represent the *physical* network whose nodes $N = \{0, 1, 2, \ldots, n\}$ correspond to the switching center (node 0) and the distribution points (nodes 1 to n) assigned to that center. Each edge (i, j) in the set E represents a current or potential feeder section between nodes i and j. This physical network does not provide a complete representation of the local access planning model since it does not differentiate the various transmission rates, transmission media, and processor types. To incorporate these modeling features, we use a multi-layer network that associates a different layer with each transmission rate. Each layer contains a copy of the original network, with additional edges that represent direct point-to-point cables. The arc flows within a layer correspond to transmission at a specific rate, and flows across layers model traffic processing. Sen, Doverspike and Dunatunga [76] propose a similar representation for inter-office planning.

Let us first consider a representation for the illustrative system summarized in table 2. This example has one service type associated with each transmission rate (DS0, DS1, and DS2) and one processor type for every pair of input–output rates. The example assumes a one-to-one correspondence between transmission rates and media, i.e., each transmission medium can accommodate only one transmission rate and every rate requires a unique medium. After developing the layered network for this simplified problem setting, we indicate network enhancements that permit us to model transmission media with overlapping frequency ranges.

The layered network, denoted as G_L, contains one layer for each transmission rate. Let us index the transmission rates, and hence the network layers and service types, from $l = 1$ to L in increasing order of frequency. Since we assume a unique medium for each transmission rate and vice versa, we can conveniently use the same index l to refer to the transmission medium that carries rate l signals. For example, in table 2 the index $l = 2$ corresponds to DS1 signals (rate 2), high speed data (service type 2), and coaxial cables (medium 2).

Each layer of G_L contains a replica of the nodes in the original (physical) network $G:(N, E)$. We denote the copy of the original node i in layer l as (i, l). The layered network contains two types of edges: *transmission edges* contained within each layer, and *processor edges* connecting different layers.

The transmission edge connecting node i to node j in layer l, denoted as edge (i, j, l), represents medium l lines connecting distribution points i and j. The flow on this edge corresponds to rate l transmission between i and j. If medium l is sectional, then layer l contains an edge (i, j, l) corresponding to each original feeder section (i, j) in the physical network G. On the other hand, if medium l is point-to-point, then layer l of G_L contains edges (i, j, l) for every pair of (original) nodes i and j that medium l can connect. Observe that the edges of this point-to-point network represent the logical, rather than physical, layout of medium l lines; physically, the medium l lines connecting distribution points i and j might use the intermediate ducts of the actual network. Transmission edges are either directed or undirected, depending on whether medium l is unidirectional or bidirectional.

The edges connecting two different layers of G_L represent traffic processors. The processor edge connecting node (i, l') in layer l' to node (i, l'') in layer l'' (with $l' < l''$) represents a traffic processor located at distribution point i that compresses rate l' traffic to rate l'' traffic. Thus, for the example of table 2, a processor edge from $(i, 1)$ to $(i, 3)$ will represent a type 3 traffic processor installed at distribution point i. Observe that processor edges are always directed from lower indexed layers to higher indexed layers since we only permit traffic compression (i.e., transformations from lower to higher transmission rates).

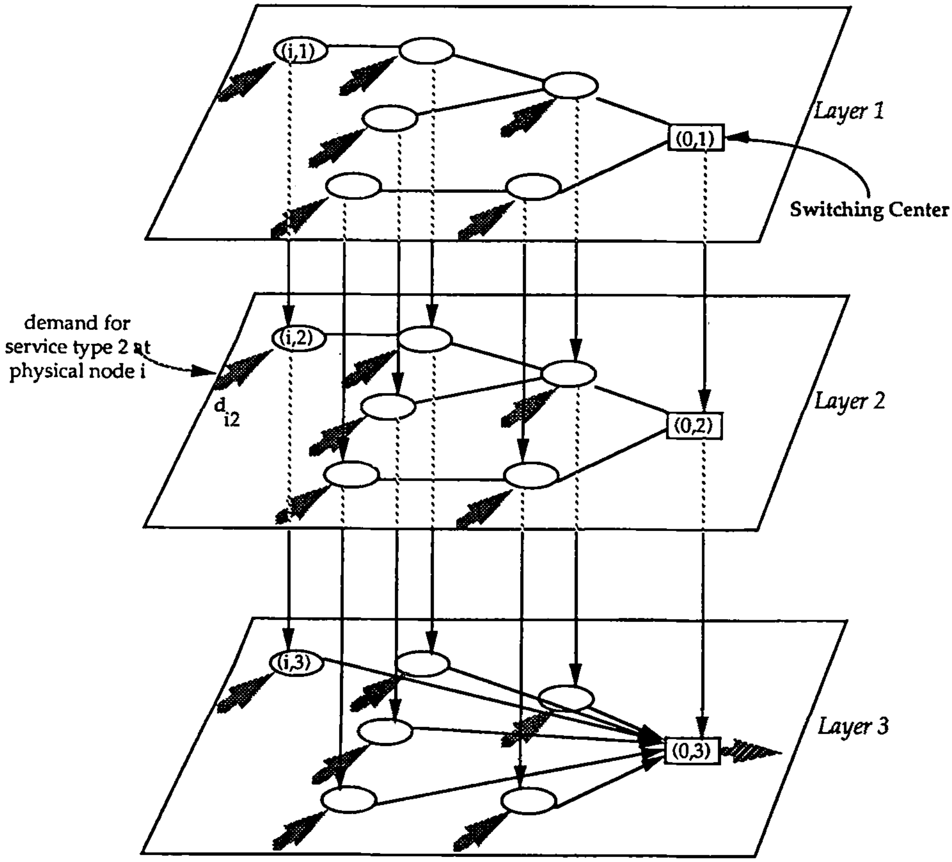


Fig. 5. Layered network representation.

Figure 5 shows the equivalent layered network representation for a local access planning problem that is defined over a (physical) network with a tree topology, and with transmission rates, transmission media, and processor types as shown in table 2. This figure assumes that twisted wire pairs and coaxial cables are sectional, while fiber cables are point-to-point from each potential fiber optic terminal to the switching center. (Thus, layers 1 and 2 have the same topology as the given physical network, while layer 3 has a star topology.)

Our objective is to represent the local access planning problem as a cost minimizing network flow formulation defined over the layered network. To complete the layered network flow formulation, we must specify: (i) the units of demand and flow measurement, (ii) the demands and supplies at different nodes, (iii) the cost functions and capacities of different arcs, and (iv) the laws governing flow conservation at each node. We consider each of these issues in turn.

3.3.1. Units of measurement

When different processor types have arbitrary traffic conversion ratios, we measure the demands, supplies, flows and capacities in each layer l in terms of the number of medium l lines (or in terms of equivalent rate l channels) required. Thus, for the example of table 2, the demand for high speed data is expressed in terms of the number of coaxial lines required for this service at each distribution point. For processor edges, we measure flows and capacities in terms of number of input lines; thus, the capacity of a type 2 processor (in table 2) is specified by the number of incoming coaxial lines. Observe that this measurement scheme implies that the units differ from layer to layer; consequently, as we elaborate later, the problem becomes a network-flow-with-gains formulation rather than a standard minimum cost network flow formulation that conserves flow at each node. In section 4.3, we consider a special situation in which a common unit of flow measurement transforms the general network-flow-with-gains problem to a flow conserving network flow formulation.

3.3.2. Demands and supplies

Each node (i, l) in layer l has "supply" d_{il} equal to the number of medium l lines required to satisfy the requirement for service type l at distribution point i. (For expositional convenience, we treat the line requirement at each node as a *supply* at that node, although we can equivalently define these requirements as node *demands*.) The switching center node $(0, L)$ in the highest rate layer L serves as the sink (or destination) node for all flows; other switching center nodes $(0, l)$, for $l < L$, serve merely as transshipment nodes.

3.3.3. Edge cost functions and capacities

To represent costs and capacities, we distinguish between three types of arcs: (i) those modeling existing facilities, (ii) those modeling additions to existing capacities or new facilities, and (iii) those modeling the collection of traffic at the switching center. To model an existing transmission facility, say, B medium l lines between nodes i and j, we introduce an edge from node i to j in layer l with zero cost but a flow capacity of B. Similarly, capacitated inter-layer edges with zero costs model existing processors. We represent additional or new transmission and processing facilities by uncapacitated expansion edges (parallel to the capacitated edges representing existing facilities). The cost of flow on each expansion or new facility edge corresponds to the *separable* cost component for transmission or processor expansion and installation. In addition, as we discussed in section 3.2, flows might incur *joint* transmission costs that depend on the combination of flows on transmission edges (in different layers) that use the same physical feeder section(s). Finally, the cost parameters to use for the processor edges $(0, l', l'')$, for $l' < l''$, incident to the switching center nodes depend on our assumption regarding permissible rates for traffic entering the switching center. In particular, if the switching center has the capability to receive multiple signal rates, each of these edges has zero cost. If, however, all entering traffic must have the same transmission rate L, the processor edge $(0, l', l'')$ carries the cost corresponding to an l'-to-l'' processor located at the switching center.

3.3.4. *Flow conservation law*

As mentioned previously, because the unit of measurement differs from layer to layer, and flows from different layers interact via the processor edges, we require a network-flow-with-gains equation to relate the incoming and outgoing flows at each node in every layer. To illustrate this constraint, consider some node $(i, 2)$ in layer 2 of fig. 5. This node has three types of incident edges:

(i) *transmission* edges $(j, i, 2)$ and $(i, j, 2)$ in layer 2 representing coaxial cables carrying DS1 traffic to and from distribution point i;

(ii) an *incoming processor* edge $(i, 1, 2)$ representing a type 1 processor located at distribution point i that compresses DS0 signals to DS1 signals; and

(ii) an *outgoing processor* edge $(i, 2, 3)$ representing a type 2 (DS1-to-DS2) processor located at distribution point i.

In addition, distribution point i has d_{i2} incoming DS1 signals corresponding to service type 2 (high speed data) at the location. Each node's incident edges (in the layered network) have associated flow variables which we define as:

$X1_i$ = throughput of the type 1 processor (expressed in terms of the number of incoming DS0 lines) located at node i;

$X2_i$ = throughput of the type 2 processor located at node i (number of DS1 input lines);

Y_{ji2} = incoming traffic (number of DS2 lines) *from* j *to* i on coaxial cables; and

Y_{ij2} = outgoing traffic (number of DS2 lines) *from* i *to* j on coaxial cables.

Observe that the line requirement variables Y_{ij2} and Y_{ji2} are "directed" even though the edge $(i, j, 2)$ is undirected (assuming coaxial cables are bidirectional). $X1_i$, $X2_i$, Y_{ij2}, and Y_{ji2} are the decision variables in the model; they determine how much transmission and processor capacity to add at location i.

The flow conservation law equates the total inflow of traffic at each node of the layered network to the total outflow at that node. First, consider node $(i, 2)$'s inflow from the type 1 processor located at distribution point i. Since this processor's throughput $X1_i$ is measured in number of input (medium 1) lines, we must translate this throughput into an equivalent number of medium 2 lines to maintain consistent units of measurement at node $(i, 2)$. The conversion ratio, call it ρ_1, of the type 1 processor enables us to express $X1_i$ in equivalent units of medium 2 lines; in particular, the number of medium 2 lines of compressed traffic arriving at node $(i, 2)$ via the type 1 processor is $X1_i/\rho_1$ (recall that conversion ratio ρ_1 is the number of input lines per output line for a type 1 processor). Thus, the input–output flow equation at node $(i, 2)$ has the following form:

$$\sum_j Y_{ji2} + X1_i/\rho_1 + d_{il} = \sum_j Y_{ij2} + X2_i. \tag{1}$$

To simplify our discussion, we have formulated the flow equation (1) for a specific example, namely, node i in layer 2 corresponding to the transmission rates, media, and processor types shown in table 2. However, the equation readily extends to more general settings in which the node in layer l receives outputs of several processor types from lower rate layers $l' < l$, and provides inputs to several processor types that compress layer l traffic to higher rates $l'' > l$.

Observe that equation (1) differs from the flow conservation equation in a standard minimum cost network flow formulation because of the "loss" factors ρ corresponding to traffic processors. Thus, to model second generation local access networks with remote electronic devices, we require a network-flow-with-gains formulation. In section 4.3, we show that, under certain assumptions regarding the traffic conversion ratios for different processor types, we can define all decision variables in terms of equivalent base rate channels instead of number of lines for different media, enabling us to use a standard network flow problem formulation that conserves flow at each node.

3.3.5. Extensions

The layered network representation extends to more complex problem settings in which each transmission medium can handle several transmission rates (rather than the single rate we assumed in table 2) and where the frequency ranges for different media overlap. To represent these overlapping ranges, we could either (i) retain the correspondence between layers and transmission rates, but introduce parallel edges within each layer to model alternative transmission media, or (ii) decompose the problem further by associating a different layer for each transmission rate–medium combination. The first representation is appropriate when we can directly interface two different media at intermediate distribution points without any additional, expensive equipment. For situations requiring interface equipment, the second representation enables us to model the cost of this equipment as the cost on interlayer edges that connect two layers corresponding to the same transmission rate but different media. In this case, we also assume that each traffic processing device requires a specific rate–medium combination for its input and output. Depending on the application context, we might use a mix of these two representations. The layered network can also model different processor technologies as parallel interlayer edges.

Within the framework of this layered network representation, the general local access network planning problem becomes:

Minimize total separable + joint transmission and processor expansion or installation cost

subject to network flow with gains constraints (1) for all nodes (i, l) of the layered network,
capacity constraints for arcs (i, j, l) or (i, l, l') representing existing transmission or processing facilities, and
non-negativity constraints for all transmission and processor throughput variables (X and Y).

Observe that this 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, the formulation permits *bifurcated* routes, i.e. two customers connected to the same distribution point might communicate with the switching center via different routes and processing steps. To prevent 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 might contain additional variables and/or constraints. For instance, we can model a fixed plus variable processor cost structure similar to fig. 3(b) by introducing additional binary variables Zm_i, denoting whether ($Zm_i = 1$) or not ($Zm_i = 0$) we install a new type m processor at node i. The fixed cost of the processor serves as the objective function coefficient for variable Zm_i, and the formulation contains additional "forcing" constraints to relate this location variable to the processor throughput variable Xm_i (see, for instance, Nemhauser and Wolsey [70], p. 7). Similarly, the formulation might contain additional constraints to model certain policy restrictions (e.g. proximity rules and contiguity requirements (see section 4.3)).

The general network flow formulation with gains and non-separable edge cost functions is difficult to solve, especially since the problem dimensions are very large for practical local access networks. Therefore, models proposed in the literature make several additional assumptions, some motivated by operational convenience and others to facilitate algorithmic development. These assumptions lead to special problem structures that solution algorithms can exploit. Next, we discuss several categories of possible assumptions.

3.4. POSSIBLE MODELING ASSUMPTIONS

The problem formulation described in section 3.3 defines a basic framework for various 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) They make the model computationally tractable. For instance, certain NP-hard location problems defined over general networks become polynomially solvable for tree networks.

(ii) They reflect uncertainties in the technology.

(iii) They reflect differences in corporate policies and practices. For instance, some telephone companies emphasize non-bifurcated routing to reduce the burden of managing/rearranging the network.

The following six areas cover various types of common modeling assumptions. Selecting different combinations of these assumptions gives rise to different models.

(a) *New versus expansion projects*: Some models apply only to designing new networks (with no existing capacities), while others apply to expansion planning as well, i.e., the latter models account for existing switching and transmission resources. As mentioned in section 3.3, to model existing capacities we require an additional set of (capacitated) parallel edges, and associated decision variables and constraints that make the model more difficult to solve.

(b) *Network topology*: Some models assume that the physical network has a tree structure. This assumption reduces the computational effort because tree networks have a unique path from each distribution point to the switching center (see section 4.3 for a tree network model that is solvable in polynomial time). Many models assume that all compressed traffic requires point-to-point media connecting each traffic processor directly to the switching center. Effectively, this assumption implies that the network layer corresponding to compressed traffic has a star topology, with the switching center as the central node.

(c) *Backfeed versus unidirectional flows*: Some tree network models incorporate *backfeed*, i.e., traffic movement away from the switching center, while others assume that all flows must be directed toward the switching center. Without backfeed, a processor that is located at node i can serve only upstream distribution points; this restriction limits the set of designs that the solution algorithm must consider.

(d) *Processor and transmission cost functions*: The generic processor and transmission cost functions described in section 3.3 can be specialized in various ways. First, all the models we discuss in section 4 ignore *joint* costs, i.e., costs that are shared by two or more transmission media. Further, if all *separable* costs are purely variable and linear functions of capacity, we can apply a network flow model (possibly with gains) to solve the local network planning problem. When we include fixed charges for cable and processor expansion, the problem becomes much more difficult to solve.

(e) *Routing restrictions*: Some telephone companies require non-bifurcated routing, i.e., all the traffic from each distribution point must follow the same route (i.e., use the same feeder sections, the same transmission medium on each section, and the same processors at intermediate distribution points). This policy facilitates network management and maintenance. Another possible routing restriction might specify that, if a node contains a traffic processor, then all entering traffic must be compressed at that node.

(f) *Single versus multiple processing steps*: Our formulation of section 3.3 permits multiple traffic processing steps in sequence. If we specify that the traffic from each distribution point must be processed at most (or exactly) once, the number of possible *homing patterns* (i.e., assignment of traffic from various nodes to processors) decreases significantly, thus reducing model complexity. Most existing second generation local access networks employ only one level of traffic processing.

These six dimensions differentiate various local access network planning models. For example, the tree location model described in section 4.3 (i) assumes that the physical network has a tree structure, (ii) permits backfeed, (iii) assumes that all enhanced (high frequency) media are point-to-point to the switching center, (iv) prohibits bifurcated routing, and (v) considers at most one processing step for traffic originating at each node. By selecting different combinations of assumptions in this manner, we can generate a diverse set of models. In the next section, we outline various modeling approaches for the single-period local network planning problem, and differentiate them using this framework.

4. Modeling approaches for local access network planning

This section describes some modeling approaches for single-period local access network planning. For each approach, we use the framework of section 3 to identify the assumptions that differentiate it from other models. Section 4.1 reviews existing models, while sections 4.2 and 4.3 cover two alternative approaches. Section 4.1 considers two models from centralized teleprocessing design, and two models proposed specifically for local access network planning. In section 4.2, we specialize the layered network model of section 3 to a fixed-charge network design problem. Section 4.3 describes a more restrictive model that applies only to tree networks; this model can be solved efficiently when designing a new network. Table 3 summarizes the differences between the various modeling approaches discussed in this section.

With a few exceptions, all of the models reviewed in this section make the following set of assumptions:

Table 3

Comparison of local access network planning models

Model type	Topology of physical network	New or existing networks?	Cost structure	Restrictions	Features	Solution method
Centralized teleprocessing design model				Point-to-point connection from concentrator to switching center; single service type		
• Capacitated concentrator location	General	New	Fixed only	No bifurcated routing	Selected design has double star topology	Optimization-based and local improvement heuristics
• Terminal layout	General	New	Fixed only	Only a single cable size can be installed	Terminal layout model usually applied with fixed concentrator location	Optimization-based approaches
Switching center connection model (Luna et al. [59])	General	New	Fixed only	Prespecified service sections, potential concentrator sites in each section, and routing from distribution points to concentrator sites; single service type	Can incorporate proximity restrictions; exactly one concentrator site selected for each service section	Shortest path/local improvement heuristic
Tree network model without backfeed (Helme et al. [43])	Tree	Both	Fixed plus variable	No backfeed	Can accommodate multiple processors in series	Dynamic programming
Network design model with layered network	General	Both	Fixed plus variable	Fixed and (linear) variable processor and cable costs; Conversion ratios for different processor types are compatible	Can accommodate multiple customer services and multiple processors in series; Cannot accommodate customer proximity	Dual ascent
Tree covering model	Tree	Both	Fixed plus variable	Contiguity assumptions; No bifurcated routing	Can incorporate proximity restrictions	Polynomial solvable for new networks; Lagrangian relaxation/ dynamic programming

- the network does not contain any existing facilities, i.e., the models apply only to the design of new facilities;
- all demand can be aggregated into a single service type;
- traffic from each distribution must undergo at most one level of traffic processing;
- all high frequency media (for carrying the output signals of the traffic processors) are point-to-point, i.e., the layer corresponding to compressed traffic has a star topology since each concentrator is directly connected to the switching center;
- joint costs are negligible, and the transmission and processor cost functions are separable with fixed and/or variable components.

Observe that the single-level traffic processing assumption implies that these models require at most two levels in the layered network representation. Further, when joint costs are negligible, and layer 2 has a star topology, we can simplify the representation to a single layer by adding the cost of the point-to-point high frequency medium to the concentrator cost at each site. In this case, the local access network planning task reduces to determining where to locate new processors, and how to connect all distribution points to the selected processor sites.

4.1. TELEPROCESSING DESIGN AND LOCAL ACCESS PLANNING 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 [11], Chandy and Lo [16], Chandy and Russell [17], Direlten and Donaldson [27], Kershenbaum and Boorstyn [48], Kershenbaum and Chou [49], McGregor and Shen [65], Mirzaian [68], and Tang, Woo and Bahl [81]).

The teleprocessing network design problem consists of three main components: (1) selecting the number and location of concentrators (*concentrator location*), (2) assigning each terminal to a concentrator (*terminal assignment*), and (3) determining how to connect every concentrator to its assigned terminals (*terminal layout*). Thus, teleprocessing design methods apply to local access network planning when we treat terminals as distribution points and the central computer as the switching center. Most teleprocessing network design methods proposed in the literature first determine the concentrator location and terminal assignment decisions using a single model, called the *Capacitated Concentrator Location Problem* (CCLP), that approximates the actual costs of connecting terminals to concentrators by (separable) assignment costs. Subsequently, a terminal layout method configures the terminal-to-concentrator interconnections based on the assignments suggested by the first phase. We first consider the CCLP as it relates to local access network planning.

(a) *Capacitated concentrator location models*

Given the fixed costs and capacities of new concentrators at each potential site and the terminal-to-concentrator connection (or assignment) cost, the CCLP selects concentrator locations and assigns each terminal to one of the selected concentrators in order to minimize the total concentrator and terminal assignment costs, subject to concentrator capacity constraints. The standard CCLP formulation assumes a single service type and provides for only one level of concentration. In terms of our local access network planning framework, CCLP models typically make the following additional assumptions:

(1) *Network topology*: CCLP models effectively assume a double star topology for the final network design: each terminal is directly connected to its assigned concentrator, and each concentrator has a direct connection to the central computer. The equivalent layered network representation contains two layers: the first layer is a complete network connecting every terminal to each potential concentrator, while the second layer is a star network connecting each potential concentrator site to the switching center.

(2) *Cost structure and capacities*: Most CCLP methods proposed in the literature apply only to the design of new networks, incorporate only fixed costs for each terminal-to-concentrator connection and for every concentrator (possibly including the cost of the concentrator-to-computer connection), and assume that the concentrator capacity (sometimes specialized as a degree constraint on the concentrator node) is prespecified rather than expandable.

(3) *Routing restriction*: CCLP models do not permit bifurcated routing. They can accommodate proximity restrictions by setting very high assignment costs for prohibited terminal-to-concentrator assignments.

An enhanced version of the CCLP model, called the *multilevel concentrator location problem*, designs a hierarchical structure with concentrators from one level homing on concentrators at the next higher level, and so on (see, for example, Konangi, Aidja and Dhas [51], and Schneider and Zastrow [75]).

If we associate plants with concentrators and customers with terminals, the CCLP is structurally similar to the *plant location* model (see, for example, the recent book by Francis and Mirchandani [33]). Algorithms for concentrator and plant location belong to five broad classes: optimization-based algorithms (e.g., Bahl and Tang [4], Cornuejols, Sridharan and Thizy [23], Pirkul [72], Woo and Tang [86]), polyhedral methods (e.g., Ward et al. [84], Lemke and Wong [57], Leung and Magnanti [58]), heuristic local improvement methods (e.g., Rousset and Cameron [74]), clustering techniques (e.g., Konangi et al. [51], McGregor and Shen [65], Schneider and Zastrow [75]), and geometric methods (e.g., Fisher and Hochbaum [31], Haimovich and Magnanti [41], Papadimitriou [71]).

We next describe the terminal layout model for teleprocessing design in terms of the layered network representation for local access network planning.

(b) *Terminal layout program*

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

(1) The model does not consider traffic compression; effectively, the problem is defined over a single-layered network.

(2) The final topology must have a tree structure.

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

(4) The model incorporates only certain special types of capacity constraints that apply to multidrop lines (which are analogous to routes in the feeder network). These constraints include degree constraints, order constraints, and load constraints (see, for example, Rousset and Cameron [74]).

Esau and Williams [30] and Sharma [77], among others, have proposed local improvement methods for the terminal layout problem. The *capacitated minimal spanning tree problem*, which is a special case with concentrator degree constraints reflecting capacities, can be solved using optimal and optimization-based heuristic methods (e.g., Chandy and Lo [16], Gavish [36], Gavish and Altinkemer [37], Kershenbaum and Peng [50]) or polyhedral approaches (e.g., Hall [42], Araque, Hall and Magnanti [2]).

4.1.2. Special local access planning models

We now describe two models devised specifically for local access network planning. Both models apply only to the design of new networks. The first method is a heuristic proposed by Luna, Ziviani and Cabral [59] to solve a variant of the local access network planning problem which we call the *service section connection problem*. The second method is a dynamic programming algorithm developed by Helme, Jack and Shulman [43] for a tree network model that prohibits backfeed.

(a) *Service section connection model*

Luna et al. [59] consider the following variant of the local access network planning problem. We are given a partition of the set of distribution points into S subsets called *service sections*, each containing a number of potential concentrator sites. For each service section, we must select exactly one site that serves all distribution points within that section. Arcs of the network interconnect concentrators and the switching center, and carry a fixed and variable cost: concentrators have fixed costs that may vary by location. The planning problem consists of: (i) selecting one concentrator site from each service section, and (ii) designing a subnetwork that connects all the selected concentrator sites to the switching center, in order to minimize the total (fixed + variable) arc costs plus the concentrator costs. Unlike the CCLP, the service section connection problem explicitly considers the topological design decisions for connecting concentrators to the switching center. However, the model ignores distribution point-to-concentrator interconnections within each service section. The model considers only a single service type, a single concentrator type and transmission medium, and does not model economies of scale in concentrator and transmission costs.

Since the service section connection model considers only design decisions for compressed traffic, we can use a single-layered network representation with the following additions to the physical network:

> For each service section, we add a *super* node whose demand equals the sum of the demands for all distribution points in that service section. We connect each super node to every potential concentrator site in the corresponding section; the fixed cost of this edge equals the cost of the concentrator. All the edges of the given physical network (interconnecting potential concentrator sites and the switching center) carry the fixed and variable costs specified in the original problem; all original nodes (i.e., potential concentrator sites) serve as transshipment points. The switching center node has supply equal to the total demand in all service sections.

Since the model considers only new facilities without any capacity constraints, an optimal solution assigns each super node to exactly one concentrator site in the corresponding service section; thus, we do not need explicit constraints to prevent the model from selecting multiple concentrator sites in each service section. Also, the optimal topology of the service section interconnection network has a tree structure, and does not bifurcate traffic flow. Because of the fixed-plus-variable cost structure, this model is a special case of the fixed-charge network design model that we present in section 4.2. Luna et al. formulate the service section connection problem as a mixed integer program, and propose a heuristic method to solve it. The heuristic starts by selecting, for each service section, the distribution point that is closest to the switching center, and finds 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 authors report computational results for three problems ranging in size from 18 nodes, 54 arcs, and 7 service sections, to 263 nodes, 752 arcs, and 117 service sections.

(b) *Tree network model without backfeed*

Helme et al. [43] propose a dynamic programming method to design new local access networks with multiple processors in series, but with a single transmission medium. The model assumes that the given network has a tree structure, does not permit backfeed, and does not account for economies of scale. Each processor has a fixed cost; transmission facilities have only variable costs. The solution 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 propose a Drop/Add heuristic for a more general local access network planning model that considers 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 processor costs. The method ignores fixed costs for cable expansion; it assumes that cable costs are directly proportional to the number of (additional) cables required. The Drop heuristic starts with an initial design containing all processor types at each node, and successively eliminates processors to reduce the total cost. A shortest path algorithm computes the total transmission cost for each candidate processor configuration.

4.2. NETWORK DESIGN MODEL

When specialized, the general layered network framework introduced in section 3 becomes the following *fixed-charge network design problem*: Given the demand between various origin–destination pairs and fixed and variable costs for each arc of a network, select a subset of arcs and route the various origin-to-destination flows (subject to conservation of flow, without gains or losses, at each node) over the selected arcs in order to minimize the total fixed plus variable arc costs. The capacitated version of this problem accounts for arc capacity constraints as well. The network design problem arises in a variety of distribution planning, manufacturing, and telecommunication contexts. It generalizes several well-known optimization models, including the plant location, shortest path, Steiner tree, traveling salesman, and minimal spanning tree problems. Magnanti and Wong [60] and Minoux [67] describe various applications and solution methods for the network design model.

To transform the general network flow-with-gains formulation of section 3 to a fixed-charge network design model, we make two assumptions:

(1) *Transmission and processor cost structures*: The network design model ignores joint costs between various transmission media and assumes, as in the example of table 2, a one-to-one correspondence between transmission rates and media, i.e., the planner preselects a preferred transmission medium for each transmission rate. We also assume that all processor and cable installation/expansion costs are piecewise linear, consisting of fixed and variable components.

(2) *Conversion ratios* for different processor types: Recall that each traffic processor requires a certain input rate (or frequency), transmits output at a higher rate, and has a specified conversion ratio ρ (defined as the ratio of input to output lines). To formulate the network design model, we assume 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 we can compress rate l_a traffic to rate l_b, either by employing a type 1 and type 2 processor in series (i.e., the type 2 processor has the same input rate as type 1's output rate), or by using a single type 3 processor (with input rate l_a and output rate l_b). The *conversion ratio*

compatibility assumption imposes the following condition on the three conversion ratios:

$$\rho_3 = \rho_1 * \rho_2,$$

where ρ_m denotes the conversion ratio for a type m processor. In other words, the messages on x lines in the rate l_a medium always require exactly (x/ρ_3) lines in the rate l_b medium, regardless of whether the compression was achieved using a type 1 and type 2 processor in tandem, or a single type 3 processor. Note that the example of table 2 does not satisfy this assumption.

Effectively, the conversion ratio compatibility assumption permits us to associate a single *conversion factor*, call it δ_l, with each layer l of the multi-layer network. The factor δ_l is defined as the number of *base rate* (e.g., DS0) channels that each type l line can accommodate. Thus, we can measure traffic in every layer in terms of the *number of equivalent base rate channels* (rather than the number of lines of the corresponding medium). This common traffic measurement unit preserves conservation of flow at each node, enabling us to transform the network-flow-with-gains equation (1) to a standard network flow conservation constraint.

Apart from these two assumptions regarding cost structures and conversion ratios, the network design model incorporates all other features of the general layered network framework described in section 3.3. In particular, it can handle general network topologies, multiple service types (as long as these do not impose unique processing requirements), sectional and point-to-point 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 cost parameters for the uncapacitated fixed charge network design 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 general piecewise linear cost functions and existing capacities.

The transmission and processing costs appear as the following edge cost functions in the layered network. 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 variable cost (per unit capacity). Thus, for a processor with a capacity of X (base rate) units, the total cost is $H_{il'l''} + v_{il'l''} * X$. We associate these fixed and variable costs with the processor edge (i, l', l'') connecting layers l' and l''. (Note that a per unit cost of $v_{il'l''}$ for the processor located at node i implies that the processor cost per input (medium l) line is $\delta_l * v_{il'l''}$.) Similarly, let F_{ijl} and c_{ijl} represent, respectively, the fixed and per unit costs for a medium l connection from node i to node j. (Again, a per unit cost of c_{ijl} implies that each additional medium l *line* from i to j costs $\delta_l * c_{ijl}$.) These two transmission cost parameters define the fixed and variable costs for the transmission edge (i, j, l).

With this set of model parameters, the uncapacitated network design solution that conserves flow at each node and 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 the model permits multiple processors in series.

We can enrich this network design model in various ways. For instance, to model transmission or processor economies of scale using piecewise linear, concave cost functions as shown in fig. 6, we introduce parallel transmission or processor edges, one corresponding

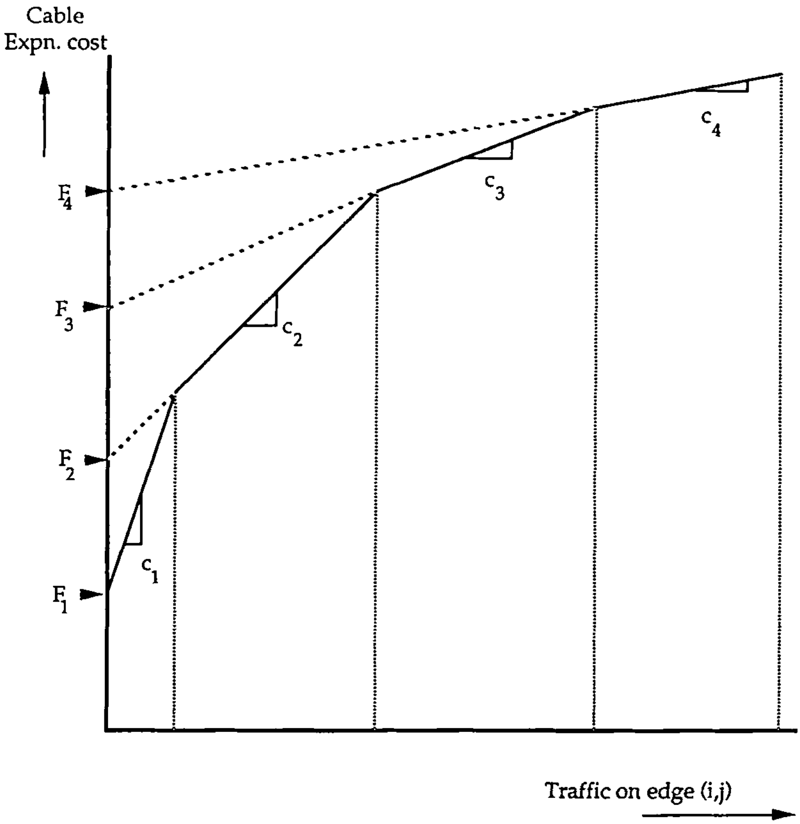


Fig. 6. Cable expansion cost with economies of scale.

to each cost range in the piecewise linear function. The rth parallel arc carries a fixed charge of F_r (which is the intercept of the rth line segment in fig. 6) and a variable cost of c_r (the slope of the rth segment). Because of concavity, the optimal solution will automatically select the appropriate parallel arc corresponding to the actual traffic, without explicit capacity constraints to define the cost ranges.

Certain properties of the optimal *uncapacitated* network design solution have special significance for local access network planning. For instance, the network design problem has an optimal solution in which each node (i, l) either processes all incoming traffic or routes all the traffic to another node on the same level. Thus, we cannot have both type l traffic leaving node i and an 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 containing, say, a 1-to-l processor. Then, this processor compresses the traffic from all intermediate nodes between i and j (inclusive). As a corollary, the problem has an optimal solution with non-bifurcated routing and satisfies the so-called contiguity property (see section 4.3).

Modeling networks with existing processor and transmission capacities, or with piecewise linear but non-concave cost functions, require explicit capacities on the edges of the layered network. In particular, suppose the network already contains X type l lines connection nodes i and j. To represent this capacity, we add a parallel arc connecting nodes (i, l) and (j, l) (in layer l) with zero fixed and variable costs, but a capacity of $X * \delta_l$ (basic traffic) units. Similarly, suppose the cost for expanding the type l transmission facilities in section (i, j) has a general (non-concave) piecewise linear structure as shown in fig. 7(a). Figure 7(b) shows the equivalent network representation with parallel arcs and appropriate arc fixed costs, variable costs, and capacities.

The network design problem and several of its variants are known to be NP-hard (Johnson, Lenstra and Rinnooy Kan [46]). Several authors have proposed heuristic- and optimization-based methods for solving the problem (for example, Billheimer and Gray [9], Boffey and Hinxman [10], Boyce, Farhi and Weischedel [12], Dionne and Florian [26], Hoang [44]). Balakrishnan, Magnanti and Wong [5] propose a dual ascent method that generates provably near-optimal solutions to the uncapacitated network design problem. Capacitated network design problems are much more difficult to solve than the uncapacitated version. Previous experience with decomposition approaches for the capacitated plant location problem, the capacitated minimal spanning tree problem, and other related models suggests that the addition of arc capacities significantly increases computational difficulty and impairs the effectiveness of these algorithms.

One of the main limitations of the network design model is its size. 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. Modeling a problem with 20 distribution points, 5 processor types, and 3 transmission media requires a network with 63 nodes and over 600 (undirected) edges. These problem dimensions probably represent the largest size that current optimization-based network design algorithms can solve within a reasonable amount of computational time (see Magnanti and Wong [60], and Minoux [67]). The next section describes a specialized model that is more tractable, since it assumes a tree network and imposes certain restrictions on distribution point-to-concentrator assignments.

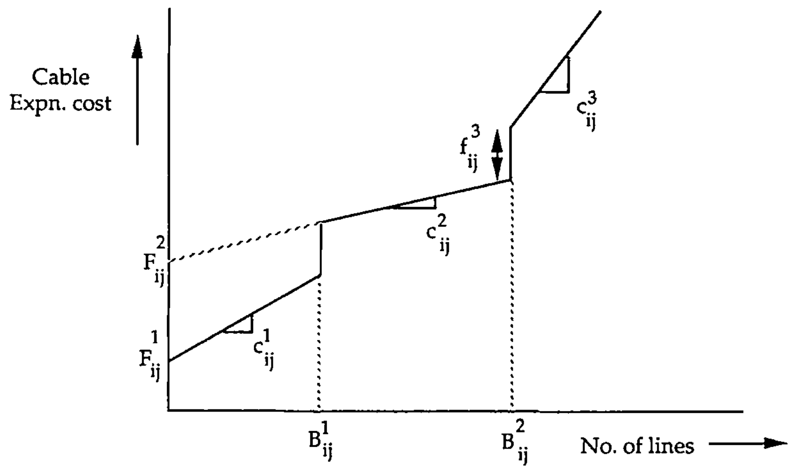


Fig. 7(a). General cable expansion cost function.

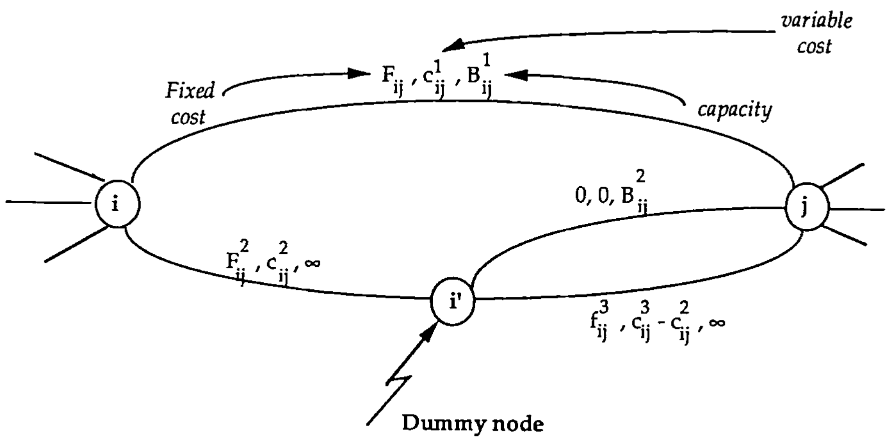


Fig. 7(b). Equivalent network representation.

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 network defining the permissible interconnections is a tree. It also makes some additional assumptions regarding the cost structure and routing policy. The model permits backfeed and can incorporate economies of scale. We first describe the model as it applies to the design of new networks, and subsequently describe an enhancement to account for existing capacities.

We say that a node i *homes* on another node j if the traffic from distribution point 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 given physical network, containing all permissible cable sections, has a tree structure, rooted at the switching center. Thus, 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 nodes on the (unique) path from i to j also home on node j. 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 to the switching center (i.e., must use the same links and undergo processing at the same node).

(5) The model ignores joint costs between media, and assumes that all high frequency media are point-to-point to the switching center. 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 (or, more generally, a piecewise linear, concave cost function) that may vary by location. Similarly, cable installation and expansion entails a fixed and variable cost that varies by section.

(7) The model can account for additional *homing costs* when node i homes on a processor located at node j. By selectively setting these homing costs to a high value, we can prohibit homing patterns that violate proximity restrictions.

Without the additional homing costs, the tree covering model has an equivalent single-layer fixed-charge network representation with some routing restrictions. The single layer consists of the original physical tree network with the following enhancements: each node of the tree is connected by a directed arc to the switching center. This arc models the traffic processor at the node as well as the point-to-point cable connecting this processor to the switching center. Every distribution node of the network has supply equal to the projected number of channels required at that distribution point; the switching center node is the sink for all flows. The contiguity assumption translates into a routing restriction specifying that each node (except the switching center) has outgoing flow (either at the base rate or at the compressed rate) on exactly one arc.

The tree covering model's assumptions permit us to transform the local access network planning task into a problem of covering all the nodes of the original tree by subtrees. Consider a local access network solution that locates a processor at node j. 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 nodes p and q both 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 single processor located within the subtree. For each potential processor location in the subtree, we can calculate the exact node and arc throughputs, and hence the exact value of the cable expansion and concentrator costs. 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, Edmonds and Wolsey [8] for optimally covering a tree with subtrees. This method is also closely related to the p-median algorithm discussed by Kariv and Hakimi [47]. The algorithm starts from the leaves of the original tree, and recursively builds the covering solution for successively larger subtrees.

Incorporating existing cable and processor capacities considerably complicates the model and its solution. Balakrishnan, Magnanti and Wong [6, 7] describe a model and Lagrangian relaxation approach that uses the dynamic programming procedure iteratively to solve an uncapacitated design subproblem. The authors describe various formulation and algorithmic enhancements to significantly improve the method's performance, and report computational results based on some actual test networks.

Next, we outline a different method, using a shortest path algorithm, to solve a special case of tree covering when the given network is linear.

4.4. LINE NETWORK MODEL

A *line network* is a simple 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 contains only (undirected) edges of the form $(i - 1, i)$, for $i = 1, 2, \ldots, n$. Figure 8 shows this structure. Suppose we want to locate p processors on this network, including the processor at the switching center. Then, by the contiguity assumption, each processor induces a line segment containing all the nodes it serves, and the union of all p line segments covers all nodes of the network. Thus, the local access network planning problem for a line network reduces to (i) determining the number of processors (p) to locate, and (ii) optimally partitioning the line network into p segments. We can

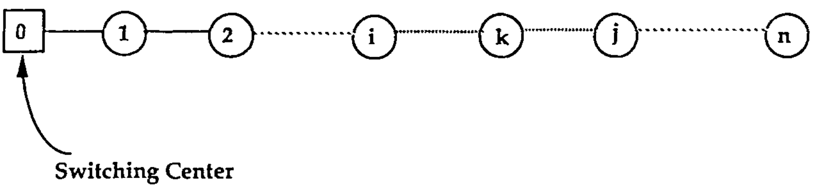


Fig. 8. Line network.

formulate this problem as a shortest path model 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. Let k_{ij} be the best processor location (i.e., the node on the line segment that minimizes total cost) for serving line segment i-to-j, and let c_{ij} be the corresponding optimal cost. Because of the line network's special structure, the costs c_{ij} can be computed efficiently.

To find the best partition of the original network, we construct an equivalent shortest path network defined over the $(n + 1)$ nodes. For every $i \leq j$, the equivalent network contains a directed arc from j to $(i - 1)$ with cost c_{ij}. Every path from node n to node 0 in the equivalent network 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. The algorithm can easily accommodate existing cable and processor capacities. For problems without backfeed, this approach can be simplified even further and is analogous to the well-known Wagner–Whitin model of production planning (Kubat [55]).

In conclusion, this section has described various models for local access network planning. We have seen several possible combinations of assumptions, each defining a separate model. Table 3 summarizes the main differences in assumptions, features, and solution methods for the models that we reviewed. The fixed-charge network design model is quite versatile, and recent advances in network design algorithms make this modeling approach computationally feasible for the design of new, medium-sized local access networks. On the other hand, the general network design model does not exploit special structures found in specific application contexts. For instance, in order to simplify the task of managing the network, some local telephone companies impose contiguity restrictions. Similarly, if each concentrator requires an umbilical (direct) connection to the switching center, ignoring shared media costs and assuming a point-to-point medium for compressed traffic might be appropriate, especially if these enhanced media can be installed in existing ducts. Making these simplifying assumptions enables analysts to use specialized algorithms, thus increasing the range of problem sizes that they can solve.

5. Concluding remarks

This paper has attempted to (i) set a backdrop for economic modeling of local access telecommunication systems by tracing the evolution of relevant technology in public telecommunication networks, (ii) provide a general framework for viewing planning and design models, (iii) summarize previous contributions from the literature, and (iv) describe some new modeling approaches. Several technological, economic, and social forces are fueling interest in these economic models, particularly the rising demand for a variety of services due to the introduction of ISDN standards, the installation of digital and fiber optic technology, and mounting competitive pressures. The traditional local access network planning tools are inadequate in the current environment because the availability of electronic traffic compression devices for the feeder network has created new ways to respond to increasing demand for telecommunication services.

For our review of planning models, we focused on static (single period) models, emphasized the differences in modeling assumptions, and briefly outlined solution methods. As our discussion of modeling approaches has suggested, the general area of local access network planning continues to provide many challenging opportunities for modeling and algorithmic development, particularly for multiperiod and multiple-service versions of the problem. The new developments in telecommunication standards and technologies should further stimulate the development of new modeling approaches.

Some of the static models offer potential for use in multiperiod settings. For instance, we might employ a decomposition method such as Lagrangian relaxation to decompose the multiperiod problem into several single-period problems. 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 approach is to use static models to generate a final target network; we might then apply a different model to plan the time-phased evolution of the current network toward the target (see, for example, Shulman and Vachani [78]).

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 distinctions in modeling fiber optic technologies.

The continuing evolution of local access network technology and the ongoing efforts to discover new ways of utilizing this technology create a number of exciting and challenging future research opportunities. For instance, remote switches and other "intelligent" hardware, when installed in the feeder and distribution networks, can perform a number of switching center functions. In many applications, customers connected to the same exchange would be able to communicate via a nearby remote switch instead of connecting all the way to the switching center. This use of 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 might be a fruitful topic for future studies.

The enormous bandwidth of fiber optic networks creates intriguing opportunities for developing new household services: for example, video programming on demand, interactive shopping services, and home telemetry With these new services, customers might become considerably more dependent on their local telecommunication system, so that any disruption in service would be very undesirable, perhaps, comparable to a power blackout. Thus, reliability issues should assume much greater importance in planning for future networks. Economic considerations drive telecommunication companies to minimize the number of links in their systems, and so the most common current local network design is a tree configuration. This design has one serious disadvantage: the failure of a single link will disconnect the network. Topologies such as ring networks (which provide two paths between every pair of nodes) overcome this deficiency, and might become more common in future local telecommunication systems. New research is needed to assess the advantages and disadvantages of such network configurations, and to design local access networks that can offer added reliability and more resistance to failure (see, for example, Groetschel and Monma [40], Monma and Shallcross [69], Pirkul, Narasimhan and De [73]).

Acknowledgement

R.T. Wong completed this work at the Krannert Graduate School of Management, Purdue University, with partial support from the Office of Naval Research under ONR Contract No. N0000-14-86-0689.

References

[1] P. Anderson, Fiber optics for loop applications: A techno-economical analysis, *Proc. Int. Symp. on Subscriber Loops and Services* (1988), pp. 213–216.

[2] R. Araque, L.A. Hall and T.L. Magnanti, Capacitated trees, capacitated routing, and associated polyhedra, Working Paper OR-232-90, Operations Research Center, M.I.T., Cambridge (1990).

[3] AT&T Bell Laboratories, *Engineering and Operations in the Bell System*, Murray Hill, NJ (1986).

[4] L.R. Bahl and D.T. Tang, Optimization of concentrator locations in teleprocessing networks, *Proc. Symp. on Computer Communication Networks and Teletraffic* (1972).

[5] A. Balakrishnan, T.L. Magnanti and R.T. Wong, A dual ascent procedure for large-scale uncapacitated network design, Oper. Res. 37(1989)716–740.

[6] A. Balakrishnan, T.L. Magnanti and R.T. Wong, Local access telecommunication network expansion: Modeling and polyhedral characterization (1990), in preparation.

[7] A. Balakrishnan, T.L. Magnanti and R.T. Wong, A decomposition approach for expanding local access telecommunication networks (1990), in preparation.

[8] I. Barany, J. Edmonds and L.A. Wolsey, Packing and covering a tree by subtrees, Combinatorica 6(1986)221–233.

[9] J. Billheimer and P. Gray, Network design with fixed and variable cost elements, Trans. Sci. 7(1973)49–74.
[10] T.B. Boffey and A.I. Hinxman, Solving the optimal network problem, Eur. J. Oper. Res. 3(1979) 386–393.
[11] R.B. Boorstyn and H. Frank, Large-scale network topological optimization, IEEE Trans. Comm. COM-25(1977)29–47.
[12] D.E. Boyce, A. Farhi and R. Weischedel, Optimal network problem: A branch-and-bound algorithm, Environ. Planning 5(1973)519–533.
[13] B. Bulcha, L.E. Kodrich, D.B. Luber, W.J. Mitchell, M.A. Schwartz and F.N. Woomer, Feeder planning methods for digital loop carrier, The Bell Sys. Tech. J. 61(1982)2129–2141.
[14] L.H. Campbell, The evolution from narrowband to broadband customer access, *Proc. Int. Symp. on Subscriber Loops and Services* (1988), pp. 144–148.
[15] G.D. Carse, New and future technologies in the local telephone network: The Victoria system, *Proc. IEEE 1986 Int. Communications Conf.* (1986), pp. 410–412.
[16] K.M. Chandy and T. Lo, The capacitated minimum spanning tree, Networks 3(1973)173–182.
[17] K.M. Chandy and R.A. Russell, The design of multipoint linkages in a teleprocessing tree network, IEEE Trans. Computers C-21(1972)1062–1066.
[18] A.J. Ciesielka and D.C. Douglas, Electronics in the suburban and light urban loop networks, The Bell Sys. Tech. J. 59(1980)417–439.
[19] A.J. Ciesielka and N.G. Long, New technology for loops – A plan for the '80's, IEEE Trans. Comm. COM-28(1980)923–930.
[20] L. Coathup, J.-P. Poirier, D. Poirier and D. Kahn, Fiber to the home – Technology and architecture drives, *Proc. Int. Symp. on Subscriber Loops and Services* (1988), pp. 281–285.
[21] J.-P. Combot and N. Epstein, The introduction of digital switching to the local network, IEEE Trans. Comm. COM-27(1979)1056–1064.
[22] J.-P. Combot, M.S.C. Tsui and R. Weihmayer, Optimal digital switching introduction into the local network, IEEE Trans. Comm. COM-29(1981)1446–1454.
[23] G. Cornuejols, R. Sridharan and J.M. Thizy, A comparison of heuristics and relaxations for the capacitated plant location problem, Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh (1987).
[24] C.S. Dawson, R.B. Murphy and E. Wolman, A history of operations research at Bell Laboratories (1925–1975), AT&T Bell Laboratories Report, NJ (1984).
[25] R. Dettmer, The integrated services digital network: Bringing home the bits, Electronics and Power (1985)25–28.
[26] R. Dionne and M. Florian, Exact and approximate algorithms for optimal network design, Networks 9(1979)37–59.
[27] H. Direlten and R.W. Donaldson, Topological design of teleprocessing networks using linear regression clustering, IEEE Trans. Comm. COM-24(1976)1152–1159.
[28] T.R. Elken, The application of mathematical programming to loop feeder allocation, The Bell Sys. Tech. J. 59(1980)479–500.
[29] W.E. Ensdorf, M.L. Keller and C.R. Kowal, Economic considerations of fiber in the loop plant, *Proc. Int. Symp. on Subscriber Loops and Services* (1988), pp. 291–296.
[30] L.R. Esau and K.C. Williams, On teleprocessing system design: Part II, IBM Sys. J. 5(1966) 142–147.
[31] M.L. Fisher and D.S. Hochbaum, Database location in computer networks, J. ACM 27(1980) 718–735.
[32] *Fortune*, ISDN: The new telephone network (November 1988).
[33] R.L. Francis and P.B. Mirchandani (eds.), *Discrete Location Theory* (Wiley–Interscience, New York, 1990).
[34] J. Freidenfelds and C.D. McLaughlin, A heuristic branch-and-bound algorithm for telephone feeder capacity expansion, Oper. Res. 27(1979)567–582.

[35] L.F. Garbanati and J.R. Palladino, Fiber optics to the subscriber: Objectives, costs and technology assessment, *Proc. Int. Symp. on Subscriber Loops and Services* (1988), pp. 286–290.
[36] B. Gavish, Formulations and algorithms for the capacitated minimal directed tree, J. ACM 30(1983) 118–132.
[37] B. Gavish and K. Altinkemer, Parallel savings heuristic for the topological design of local access tree networks, *Proc. IEEE INFOCOM '86* (1986), pp. 130–139.
[38] A.E. Gibson and D.B. Luber, Critical section methods for loop plant allocation, The Bell Sys. Tech. J. 59(1980)99–117.
[39] J.M. Griffiths, *Local Telecommunications 2: Into the Digital Era* (P. Peregrinus, London, 1988).
[40] M. Groetschel and C.L. Monma, Integer polyhedra arising from certain network design problems with connectivity constraints, Report No. 104, Institut für Mathematik, Universität Augsburg, Augsburg, Germany (1988).
[41] M. Haimovich and T.L. Magnanti, Extremum properties of hexagonal partitioning and the uniform distribution in Euclidean location, SIAM J. Discr. Math. 1(1988)50–64.
[42] L.A. Hall, Polyhedral structure of capacitated trees and approximation methods for deterministic scheduling, Ph.D. Thesis, Operations Research Center, MIT (1989).
[43] M.P. Helme, C. Jack and A. Shulman, Planning for new services in the local loop, *Proc. Int. Telecommunication Conf.* 12 (1988), pp. 5.2B.1.1–5.2B.1.12.
[44] H.H. Hoang, A computational approach to the selection of an optimal network, Manag. Sci. 19(1973)488–498.
[45] H.H. Hoang and H.T. Lau, Optimizing the evolution of digital switching in a local telephone network, *Proc. IEEE Int. Conf. on Communications* (1984), pp. 179–184.
[46] D.S. Johnson, J.K. Lenstra and A.H.G. Rinnooy Kan, The complexity of the network design problem, Networks 8(1978)279–285.
[47] O. Kariv and S.L. Hakimi, An algorithmic approach to network location problems II: The *p*-medians, SIAM J. Appl. Math. 27(1979)539–560.
[48] A. Kershenbaum and R.B. Boorstyn, Centralized teleprocessing network design, *Proc. Nat. Telecommunications Conf.* (1975), pp. 27.11–27.14.
[49] A. Kershenbaum and W. Chou, A unified algorithm for designing multidrop teleprocessing networks, IEEE Trans. Comm. COM-22(1974)1762–1772.
[50] A. Kershenbaum and S. Peng, Neighbor finding algorithms for CMST calculation, *Proc. IEEE INFOCOM '86* (1986), pp. 140–147.
[51] V.K. Konangi, T. Aidja and C.R. Dhas, On the multilevel concentrator location problem for local access networks, *Proc. IEEE Globecom '84* (1984), pp. 912–915.
[52] W.L.G. Koontz, Economic evaluation of loop feeder relief alternatives, The Bell Sys. Tech. J. 59(1980)277–293.
[53] P. Kopp, A family of solution procedures for the routing problem in medium-term planning of mixed analog–digital transmission networks, *Proc. 3rd Int. Networks Symp.* (1986), pp. 138–142.
[54] D.J. Kostas, Transition to ISDN – An overview, IEEE Comm. Magazine 22(1984)11–17.
[55] P. Kubat, Models for allocation of remote switching units to a local area access network, Internal Report, GTE Laboratories, Inc., Waltham, MA (1985).
[56] J.S. Lavin, Optimally rehoming local serving offices to a new point of presence, AT&T Tech. J. 66(1987)50–54.
[57] P. Lemke and R.T. Wong, Characterizations and descriptions of facets for the *k*-median location problem (1990), in preparation.
[58] J. Leung and T.L. Magnanti, Valid inequalities and facets for capacitated plant location problems, Math. Progr. 44(1989)271–292.
[59] H.P.L. Luna, N. Ziviani and R.H.B. Cabral, The telephonic switching centre network problem: Formalization and computational experience, Discr. Appl. Math. 18(1987)199–210.

[60] T.L. Magnanti and R.T. Wong, Network design and transportation planning: Models and algorithms, Trans. Sci. 18(1984)1–55.

[61] L.G. Mason, Network modernization with capital budget constraints, *Network Planning Symp.*, Brighton, UK (1983), pp. 5–9.

[62] L.G. Mason, An aggregate model for network modernization with constraints, IEEE Trans. Comm. COM-32(1984)1073–1079.

[63] L.G. Mason and J.-P. Combot, Optimal modernization policies for telecommunications facilities, IEEE Trans. Comm. COM-28(1980)317–324.

[64] U. Mazzei, C. Mazzetti and G. Roso, Economics of digital carriers and fiber optic systems in subscriber loops, *Proc. Int. Symp. on Subscriber Loops and Services* (1988), pp. 133–137.

[65] P. McGregor and D. Shen, Network design: An algorithm for the access facility location problem, IEEE Trans. Comm. COM-25(1977)61–73.

[66] M. Minoux, Network synthesis and dynamic network optimization, Ann. Discr. Math. 31(1987) 283–324.

[67] M. Minoux, Network synthesis and optimum network design problems: Models, solution methods and applications, Networks 19(1989)313–360.

[68] A. Mirzaian, Lagrangian relaxation for the star–star concentrator location problem, Networks 15(1985)1–20.

[69] C.L. Monma and D.F. Shallcross, Methods for designing communication networks with certain two-connected survivability constraints, Oper. Res. 37(1989)531–541.

[70] G.L. Nemhauser and L.A. Wolsey, *Integer and Combinatorial Optimization* (Wiley–Interscience, New York, 1988).

[71] C.H. Papadimitriou, Worst case and probabilistic analysis of a geometric location problem, SIAM J. Comput. 10(1981)542–557.

[72] H. Pirkul, Location of concentrators in designing local access networks, *Proc. IEEE INFOCOM '86* (1986), pp. 148–154.

[73] H. Pirkul, S. Narasimhan and P. De, Locating concentrators for primary and secondary coverage in a computer communication network, IEEE Trans. Comm. 36(1988)450–458.

[74] G. Rousset and H. Cameron, Design and optimization of private data networks, *Proc. 3rd Int. Network Planning Symp.* (1986), pp. 50–54.

[75] G.M. Schneider and M.N. Zastrow, An algorithm for the design of multilevel concentrator networks, Computer Networks 6(1982)1–11.

[76] S. Sen, R.D. Doverspike and M.S. Dunatunga, Unified facilities optimizer, Working Paper, Systems and Industrial Engineering Department, University of Arizona, Tucson (1989).

[77] R.L. Sharma, Design of an economical multidrop network topology with capacity constraints, IEEE Trans. Comm. COM-31(1983)590–591.

[78] A. Shulman and R. Vachani, An algorithm for capacity expansion of local access networks, *Proc. IEEE INFOCOM '90* (1990), pp. 221–230.

[79] M.A. Sirbu and D.P. Reed, An optimal investment strategy model for fiber to the home, *Proc. Int. Symp. on Subscriber Loops and Services* (1988), pp. 149–155.

[80] R.K. Snelling and K.W. Kaplan, Current and future fiber optics applications – Operating telephone company perspective, *Proc. IEEE Globecom '84* (1984), pp. 604–607.

[81] D.T. Tang, L.S. Woo and L.R. Bahl, Optimization of teleprocessing networks with concentrators and multiple connected terminals, IEEE Trans. Comput. C-27(1978)594–604.

[82] *The Economist*, A survey of telecommunications: New lines for old (October, 1987), pp. 3–34.

[83] A.G. Toth, E. Colombini, P.J. MacLaren and R.K. Yates, Fiber in the local exchange network: A planning overview, *Proc. IEEE Int. Communications Conf.* (1985), pp. 520–526.

[84] J.E. Ward, R.T. Wong, P. Lemke and A. Oudjit, Properties of the tree k-median linear programming relaxations, Working Paper, Krannert Graduate School of Management, Purdue University, West Lafayette (1988).

[85] P.E. White, The broadband ISDN – The next generation telecommunications network, *Proc. IEEE 1986 Int. Communications Conf.* (1986), pp. 385–390.
[86] L.S. Woo and D.T. Tang, Optimization of teleprocessing networks with concentrators, *Proc. Nat. Telecommunications Conf.* (1973), pp. 37C1–37C5.
[87] Y. Yamamoto, H. Yamamoto and H. Oikawa, Access network design with optical fibers, *Proc. IEEE Globecom '84* (1984), pp. 637–641.