NTIA-REPORT-79-26

Switch Element Capacities in Access Area Digital Switching Systems

R. F. Linfield
M. Nesenbergs

Prepared for
U.S. Army Communications Systems Agency
Ft. Monmouth, New Jersey 07703

U.S. DEPARTMENT OF COMMERCE
Juanita M. Kreps, Secretary

Henry Geller, Assistant Secretary
for Communications and Information

September 1979

DISCLAIMER

The views, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision unless designated by other official documentation.

PREFACE

The study reported here was performed for the U.S. Army's Communications Systems Agency (CSA), Ft. Monmouth, NJ, by the Institute for Telecommunication Sciences (ITS), Boulder, CO. It is part of a continuing program in support of that agency's Access Area Digital Switching System (AADSS) program on project orders 501-RD, 804-RD, and 809-RD.

Previous related study reports are as follows:

"Parametric Cost Alternatives for Local Digital Distribution System," M. Nesenbergs and R.F. Linfield, OT Report 76-95, March 1976.

"Preliminary Evaluation of Hub Alternatives for Access Area Digital Switching," J.C. Blair, Special unpublished ITS report to CSA, October 1977.

"Access Area Switching and Signaling: Concepts, Issues, and Alternatives," R.F. Linfield and M. Nesenbergs, NTIA Report 78-2, August 1978.

"Control Signaling in a Military Switching Environment," R.F. Linfield, NTIA Report 79-13, January 1979.

Administrative and technical monitoring of this current study contract was performed by Mr. T. Michelli of CSA. Technical and management supervision of the program at ITS was provided by Dr. P.M. McManamon.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SWITCH ELEMENT CAPACITIES IN ACCESS AREA DIGITAL SWITCHING SYSTEMS

R.F. Linfield and M. Nesenbergs*

The capacity of a digital circuit switching system is defined in terms of its four major elements: the traffic offered by the interface element, the maximum traffic carried by the switch matrix, the maximum number of call attempts handled by the control processor, and the maximum number of calls handled by the signaling elements. These element capacities may be engineered for the expected traffic (e.g., call attempts and holding times), the performance objectives (e.g., blocking probabilities and tolerable delays), the processor capabilities (e.g., speeds, memory sizes, service features), and the signaling techniques (e.g., common channel or per-channel signaling).

In a properly engineered system, the interrelations between the capacity of all elements should be considered. This report discusses such interrelationships and characterizes representative switch configurations on the basis of the four major elements. The results obtained have applications in developing non-tactical networks for military access areas where the communications profile is known. This is demonstrated for Ft. Monmouth and its environs. Estimates of traffic statistics and switching requirements are made with the aid of available Ft. Monmouth terminal density profiles.

1. OBJECTIVE AND SCOPE

The Institute for Telecommunication Sciences (ITS) has been conducting a series of studies on access area communications for the U.S. Army Communications Systems Agency (CSA). The purpose of these studies is to assist the Army in its competitive concept evaluations and procurement efforts for upgrading non-tactical networks in the military access areas. This is to be accomplished by providing CSA, plus other Army or DoD organizations, with needed information on the state-of-the-art capabilities of switching, signaling and distribution systems.

Previous efforts conducted by the Institute and reported earlier include studies of cost alternatives for local digital distribution systems, pre-

*The authors are with the Institute for Telecommunication Sciences, National Telecommunications and Information Administration, U.S. Dept. of Commerce, Boulder, CO 80303.

liminary switch hub evaluations, digital switching concepts, and control signaling techniques. The study reported here extends the previous work on switching concepts to include a realistic assessment of switch system capacities. Particular emphasis is placed on the processor capabilities in stored program controlled (SPC) systems.

Digital switches with stored program control, and especially when augmented with digital transmission facilities, provide new service features and system functions to the user. It appears that this digital technology can, at the same time, permit reductions in space, power, and eventually, operating and maintenance costs.

The application of digital technology to military access area posts, camps and stations further extends the digitization process from the Defense Communication System (DCS) long-haul backbone to the local level. This helps resolve many otherwise complex interface issues.

Figure 1 is a simplified deployment diagram which illustrates how centralized switching hubs in the access area provide gateways to the DCS network and to commercial common carrier networks. The central hubs also provide all required local access area telecommunication services. Remote hubs, PABX's, multiplexers, or concentrators located at terminal cluster points assist the central hub in these services.

This report defines the parameters which determine the capacity of modern switching systems (Sec. 2). Existing commercial switching systems are used to exemplify the characteristics of representative switches (Sec. 3). Numerical parameter values are developed for a range of switch sizes. These are chosen to have application in the military access area (Sec. 4). Teletraffic engineering concepts are introduced in Section 5. The results are applied to a specific military complex whose communications profile has been previously determined (Sec. 6). The concluding paragraphs outline additional key studies which are necessary to specify the digital switching system for the access area (Sec. 7).

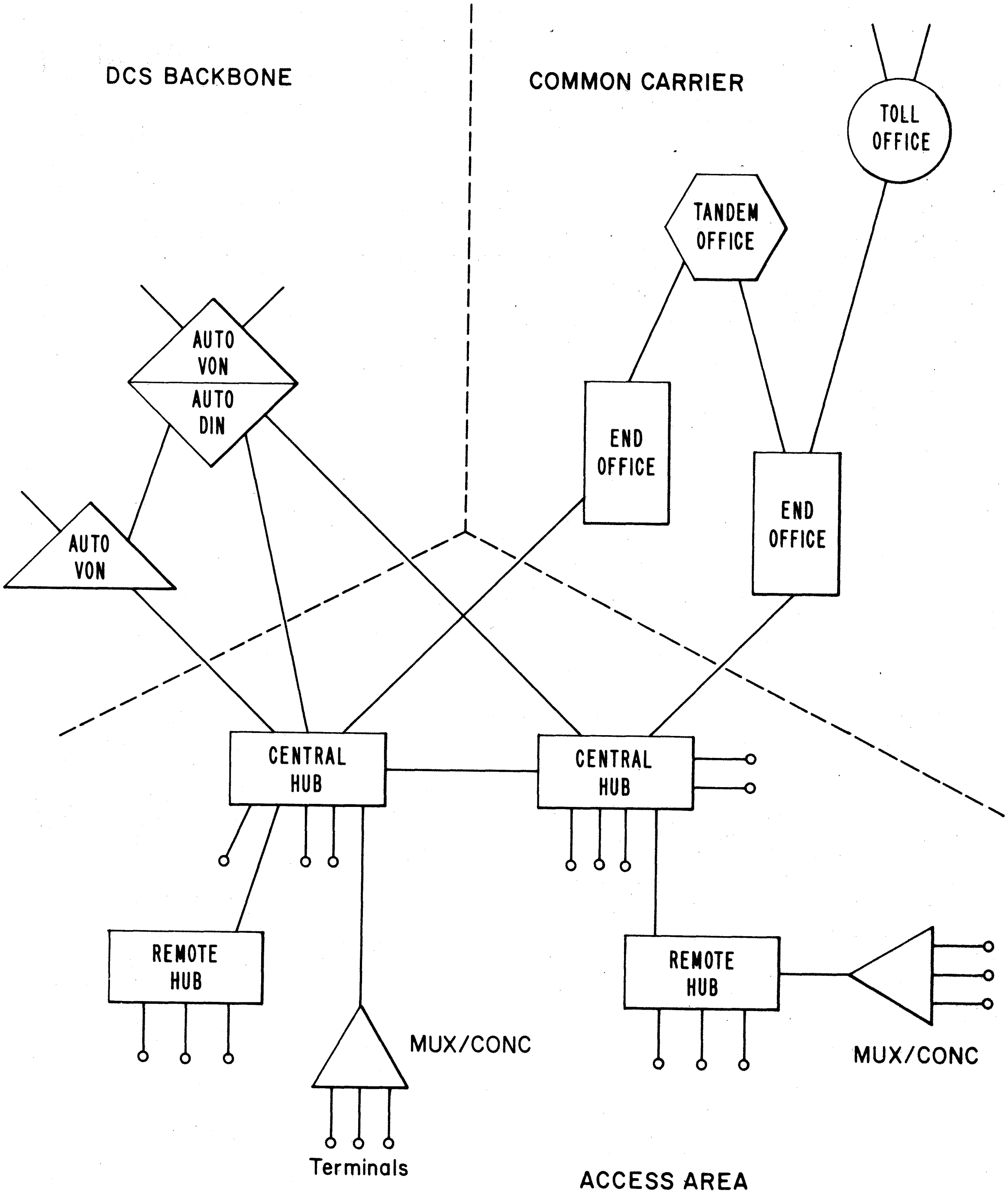


Figure 1. Basic deployment concept of access area switch hubs and long-haul networks.

2. SWITCHING CONCEPTS AND DEFINITIONS

A switch is a device used to interconnect two or more circuits or to store, process, and forward telecommunications traffic. In a telecommunications network, switching permits more economic sharing of transmission facilities between switching centers and stations on a periodic or demand basis. Switches are used either to configure the transmission paths through a network or to provide temporary access to the network on a potential message delay basis. In the first case, termimals are connected in real time. They are said to be circuit switched. In the second case, one has store-and-forward switching. It imposes a prospective penalty of potential service delays in order to use transmission facilities more efficiently.

The major elements of a network switch are depicted in Figure 2. The four basic elements and their primary functions are:

1. The interface element, which matches the switch to the transmission facilities (lines to terminals, trunks to other switches) or to other network resources (signaling channels, service circuits).
2. The configuration or access element, which can change its state in space, time or frequency, depending on control instructions.
3. The control element, which makes decisions and activates changes.
4. The signaling element, for sending and receiving control messages.

Switches can be implemented in numerous ways using a variety of technologies for each element. In some configurations, the functions performed by separate elements may overlap. The type of switch selected for a particular application depends on the nature of traffic it must handle and the transmission environment in which it must operate.

In the following subsections the technologies available to the four basic elements are defined. The evolution of switching from manually controlled switches to electronic switches with stored program control is reviewed. Criteria for selecting a certain type of switch for a given application are discussed. Finally, the call processing functions performed by each element of a circuit switch are summarized.

These introductory concepts and definitions serve as the basis for characterizing modern, commercially available, circuit switches.

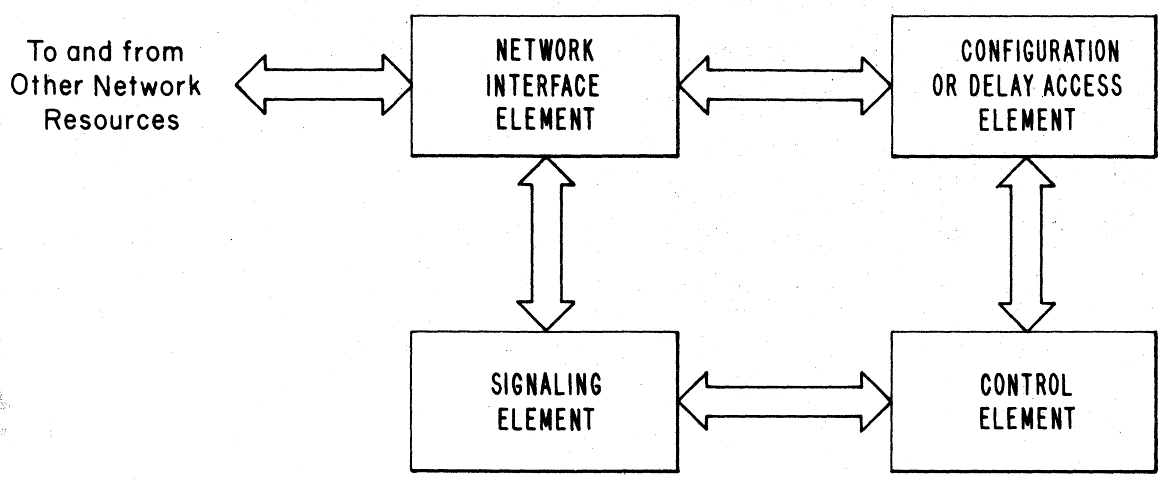


Figure 2. The four basic elements of a telecommunications network switch.

2.1. Switching System Classification

Telecommunication switch systems today fall into two basic classifications -- circuit switches and store-and-forward switches (see Fig. 3). Circuit switches have been used for switching telephone networks for nearly one hundred years. Circuit switches generally involve bidirectional information transfer with essentially no delay. Store-and-forward switches, in contrast, involve unidirectional information transfer that may be delayed. An early form of store-and-forward switch was the torn-tape message center. Today high-speed computers with mass memories perform the message reception, processing, storage and forwarding functions.

Circuit switches are normally distinguished by the technologies used in two of the basic elements, the switching matrix and the control.

Store-and-forward switches are typically categorized by the method used to format the message (e.g., length and header). The format affects the delay in the network. Distinguishing aspects of both types are discussed in the following subsections.

2.1.1. Circuit Switch Technology

The switching matrix in a circuit switch may be either analog or digital, as indicated in Figure 3. The analog and digital matrices may be subdivided further into either space-divided (SD), time-divided (TD), or frequency-divided (FD) networks. One talks accordingly of space-division, time division, or frequency division circuit switch technologies. These switching matrices are used to establish a direct connection between communication terminals.

A space-division matrix provides a physical connection between two or more circuits via a unique series of spatial points, at a given time. Time-division and frequency-division matrices provide this connection through equivalent unique points in time or in frequency.

Analog space-division switches have been commonly used in the past. Today, the trend is toward digital time-division switches. Frequency-division switches are rarely used.

In analog circuit switches, information is represented in analog form. For instance, voice signals may be modulated on carriers. Digital circuit

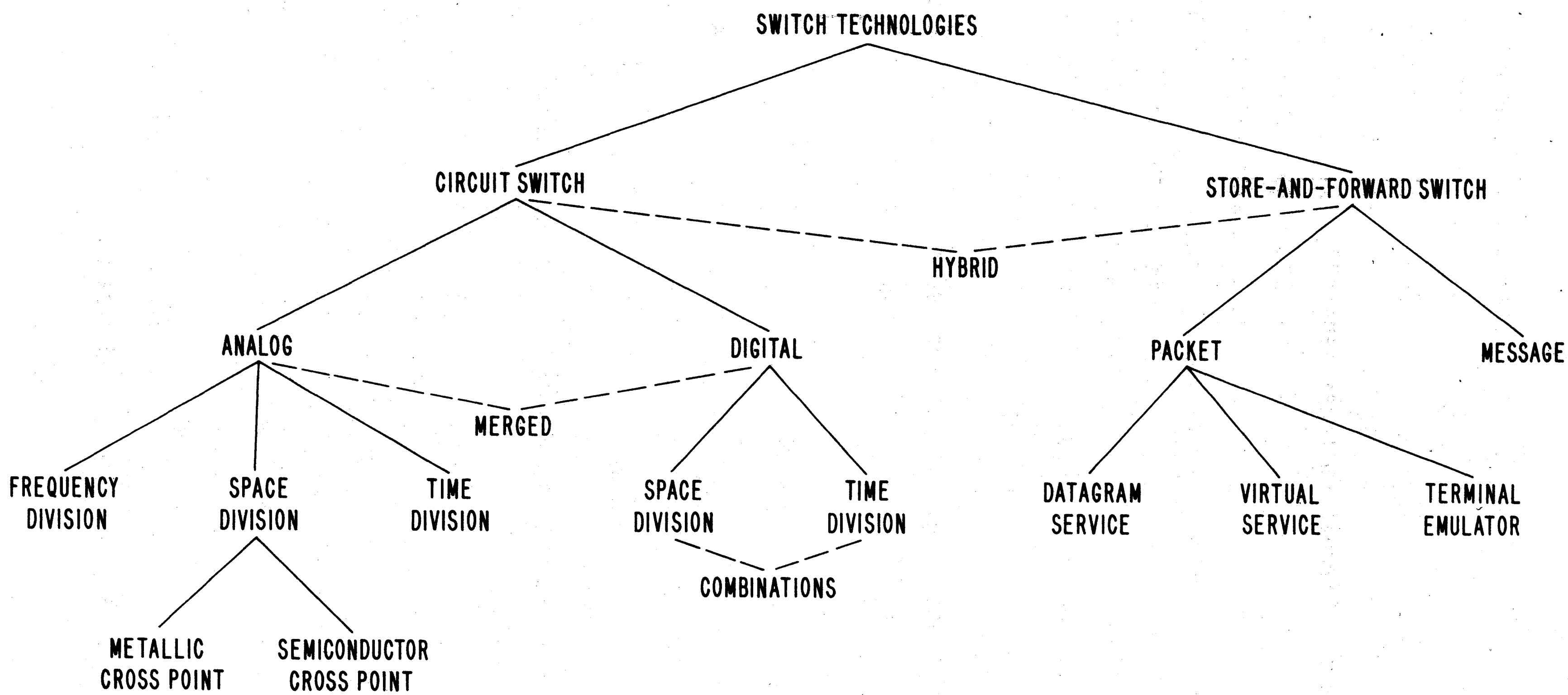


Figure 3. Switch technology classification.

switches route information in digital form. When digital switches are used in telephony, some form of voice digitization process (VDP) is required. Analog voice signals are sampled and converted to binary words. Various means of conversion are used, the most common being pulse code modulation (PCM). Subject to memory compatibilities, the binary words from many sources are interlaced serially onto a common highway or bus with the aid of a time-division multiplexer (TDM) or a concentrator. Switching is accomplished by interchanging the time slots on the bus. This time slot interchanging (TSI) process may be combined with a space-divided network. The advantage of a joint spatial and time gating technique between buses is to provide parallel sets of available time slots. There are many possible combinations of time and space division switching stages. Examples of some analog and digital switch configurations are shown in Figures 4 and 5.

In general, the analog switches use linear crosspoints in the matrix. An exception occurs when an analog signal is converted to pulse width or pulse time modulation. No confusion results if these are still considered analog switches.

The digital switch configurations shown in Figure 5 employ non-linear circuit elements. Binary bit streams could, of course, be switched by a linear crosspoint. However, conventional analog signals and pulse amplitude modulated signals cannot be switched using non-linear devices.

It is feasible to combine analog and digital technologies in a single switch. An example of such merged technology switch is the tactical switch being developed by the Joint Tactical Program Office known as TRI-TAC. This switch, designated the AN/TTC-39, uses electronic crosspoints in a space-divided matrix to handle analog terminations, and a time-divided matrix to handle digital terminations. Continuously variable slope delta (CVSD) modulation is used for the VDP between the two matrices. A simplified block diagram of the AN/TTC-39 merged technology switch is shown in Figure 6.

One aspect of switch classification which is not included in Figure 3 is the coexistence of two-wire and four-wire switching. A telephone connection can carry voice signals in both directions simultaneously over one pair of wires. This is normal practice between a telephone and a circuit switch. Switching is often, but not always, accomplished by using bidirectional

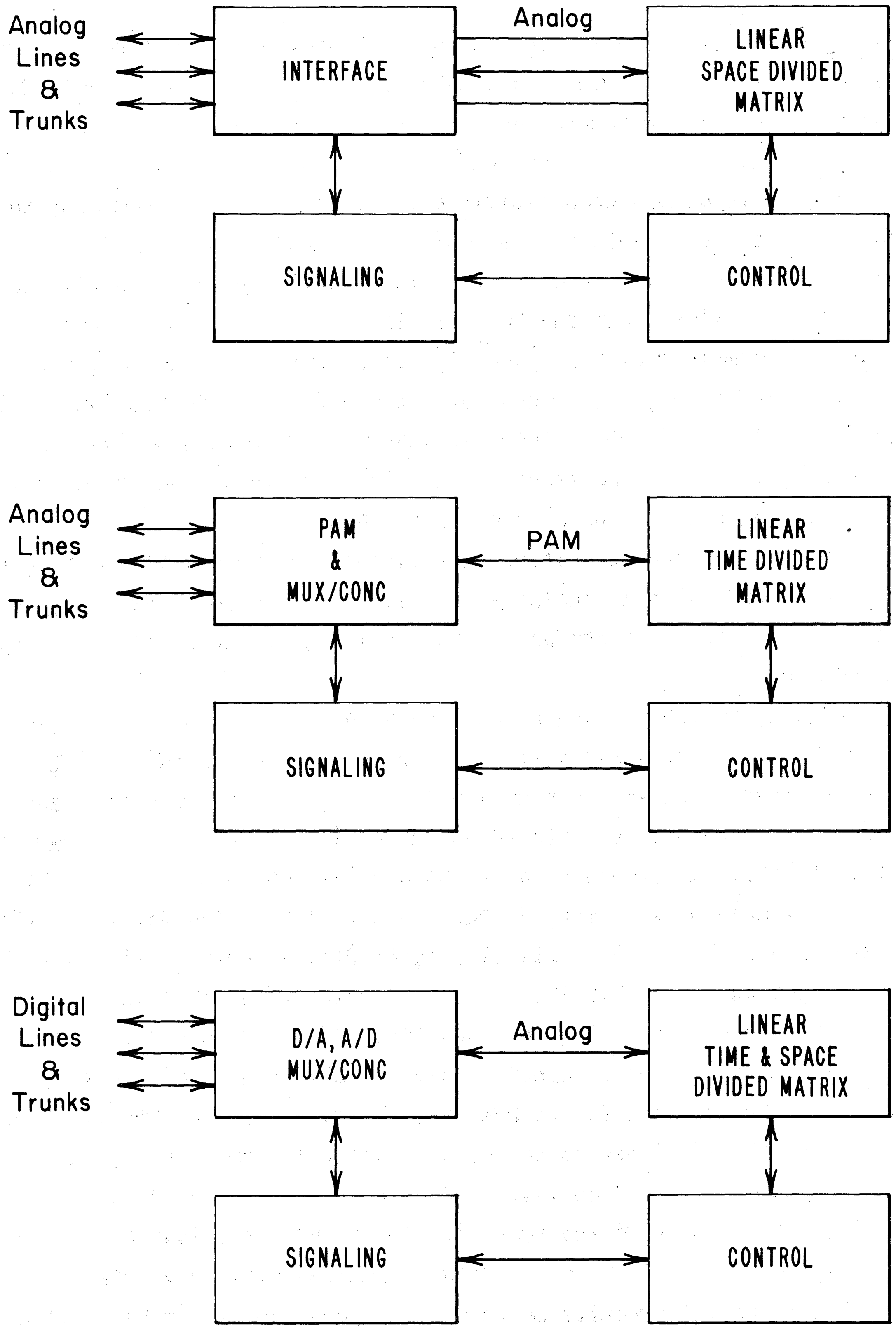


Figure 4. Analog switch configurations.

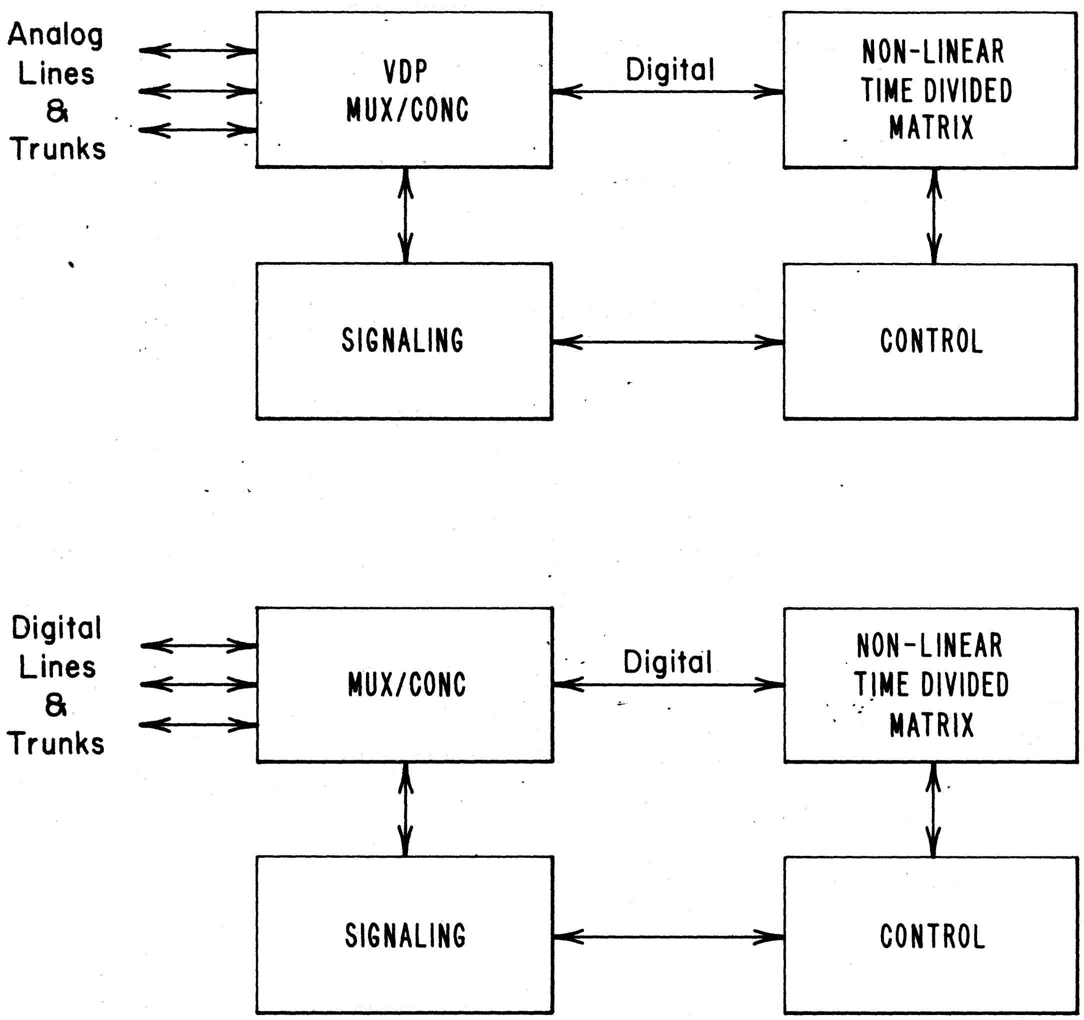


Figure 5. Digital switch configurations.

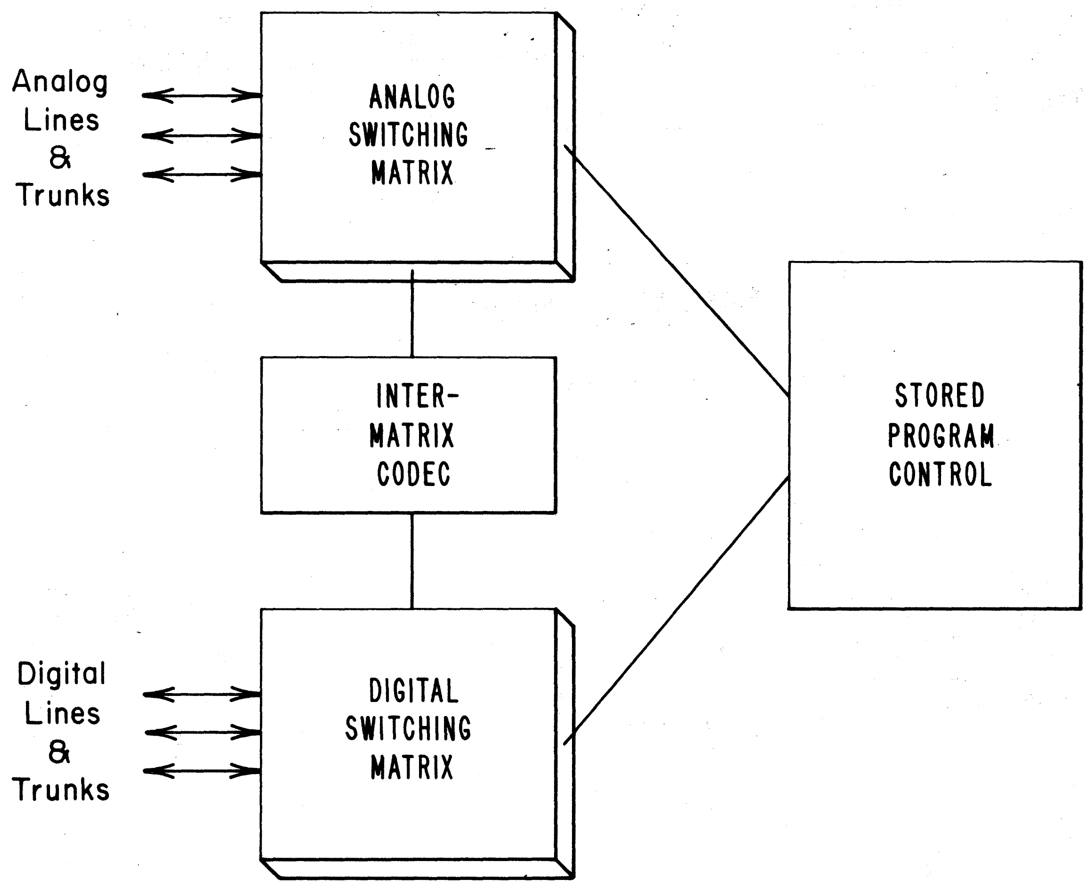


Figure 6. The AN/TTC-39 merged technology switch.

crosspoints. Cost and return-loss considerations affect the crosspoint selection. In an analog carrier system using frequency-division multiplexing (FDM) and digital carrier using time-division multiplexing (TDM), however, the two directions of transmission must be separated because amplifiers, modulators, codecs, etc., are inherently unidirectional devices. Thus within the matrix, each direction of transmission and the switching requires two pairs of wires. This four-wire transmission and switching loop can be extended to the subscribers' terminals for digital voice and data transmission. There are possibilities of easier encryption. The four-wire switching matrix has the advantage of infinite return loss. It minimizes echo and singing problems at the switching point. Digital switches considered in this report are assumed to perform four-wire switching. "Four-wire" operation implies full duplex operation with separate time-slots in the time-divided networks.

2.1.2. Control Technology

There are a variety of techniques used to control a circuit switch. The techniques depend on the functions to be performed and the matrix configuration. Figure 7 shows the various control technologies employed by circuit switching systems and the approximate year each technology was introduced.

Early circuit switches were analog in nature and used space-division matrices to interconnect transmission facilities. These switches started with manually controlled switchboards that used plugs and jacks. They evolved into electromechanical step-by-step switches with direct control, panel switches with indirect control, and crossbar switches with common relay-type control. Gas tubes and magnetic-reed relays were also used for crosspoints. In the late 1950's, electronic switches were introduced. Electronic gates were implemented mostly with semiconductor devices. They replaced not only the metallic crosspoints of the space-divided matrix, but also the wired circuitry of the common control logic. As switching systems increased in size and new service features and system functions were added, many of these wired logic control functions were replaced with software routines stored in the memory of a central processor.

Large scale integrated circuit developments have made it practical to use wired or permanently stored program logic which resides in read-only memories (ROM). This so-called firmware cannot be changed easily. It also

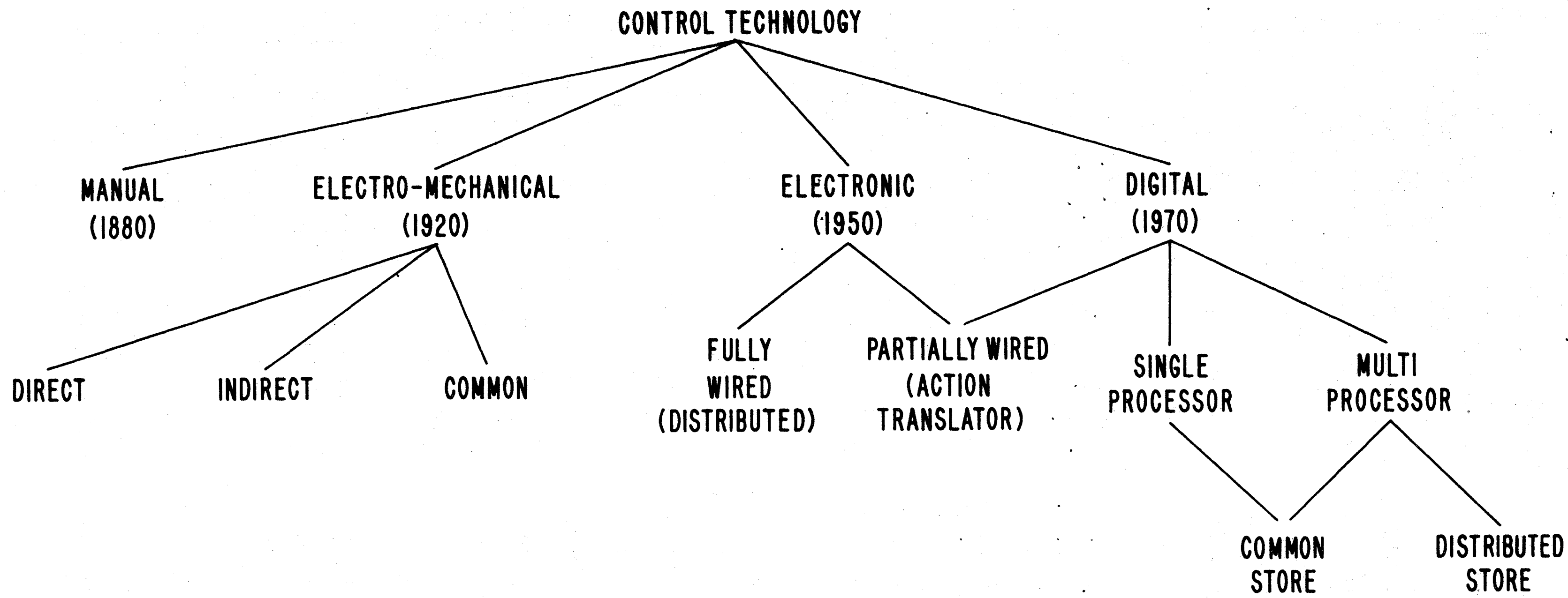


Figure 7. Control technologies used in circuit switching systems.

does not require continuous power. Sometimes the entire control logic may be wired program logic (WPL) in this form. This class of control is known as a distributed logic system. Other systems may use combinations of wired program logic and stored program control (SPC) logic which is in the form of software. Software may be electronically stored in random or structured access memories (RAM). The contents of RAM's are easily changed. Usually, programs are loaded into the RAM from an external magnetic tape or disk. This type of software requires reloading after power failures. An example of this type of WPL and SPC system is the action translator. In the trade, the term action translator is synonymous with stored program translator. It refers to memory devices that perform a specialized table look-up function. That function serves to control only the sequence of call actions (Joel, 1977).

Stored program controlled systems employ a variety of system architectures. The RAM may be associated with a single processor, with multiprocessors or several processors.

A more detailed discussion of the evolution of circuit switches and their control is given by Joel (1977) and Leaky (1977).

2.1.3. Store-and-Forward Switches

Unlike circuit switches, store-and-forward switches do not require a switching matrix for establishing a connection. Instead, the information (e.g., digital data) to be transferred is stored in a memory device until the network is ready for transmission. There are two basic types of store-and-forward switches -- message and packet (see Fig. 3). Message switches are designed to handle blocks of digital data. The block size may be variable within some maximum bound. Each block is usually headed by its own destination address. The message switch stores the first block. The block is forwarded towards the destination address when a suitable transmission facility becomes available. Thus, the message block can be delayed by varying amounts that depend on the available channels and their topology. Delay times often range from minutes to hours. The storage facilities are used in lieu of more costly transmission facilities.

Store-and-forward packet switches are similar to message switches except each individually addressed block is limited to short sequences of information bits called packets. The packet size is usually fixed. A typical length is of the order of 1000 bits. The packet consists of information bits and the header which contains source, destination, type, priority and other transmission information. No long-term storage is required in a packet switch. Packets are held only for short periods and are sent individually toward their destination as rapidly as possible. Alternate routes may be used for individual packets in a sequence of packets. Delay times vary due to the alternate routing but are typically less than one hundred milliseconds. Packets may be lost in transit or even discarded if storage facilities are exceeded. In these cases, retransmission is normally requested.

Store-and-forward switches may process digital data which resides in storage or as it flows through the switch. Communication processing functions include code and speed conversion, that make the switch compatible with a variety of terminals. Processing may also include other functions, such as error detection and correction, which assist in the transmission of information. The amount of processing provided depends to some extent on the users' requirements and the user interface. Three types of interfaces have been defined by Roberts (1977) for packet switched networks. They are: the datagram interface, the virtual circuit interface, and the terminal emulation interface.

In many instances, the user can select the interface of his choice. The service features provided to the user by these interfaces are functions of the amount of processing provided. The datagram interface requires the least amount of processing. Datagram terminals, not the interface, are responsible for detecting duplicate and lost packets, for requesting retransmissions, for sequencing packets, and for overall flow control through the network. With virtual circuit terminals, the switch interface provides most of these functions. However, some functions, such as flow control, are still the responsibility of the host computer or its front end processor. In the terminal emulation interface, all of the functions are the responsibility of the packet switching network. Almost any type of host computer is therefore directly compatible with any communications process.

The basic building blocks of a circuit switch and a store-and-forward switch are compared in Figure 8. The control logic for both switches is a

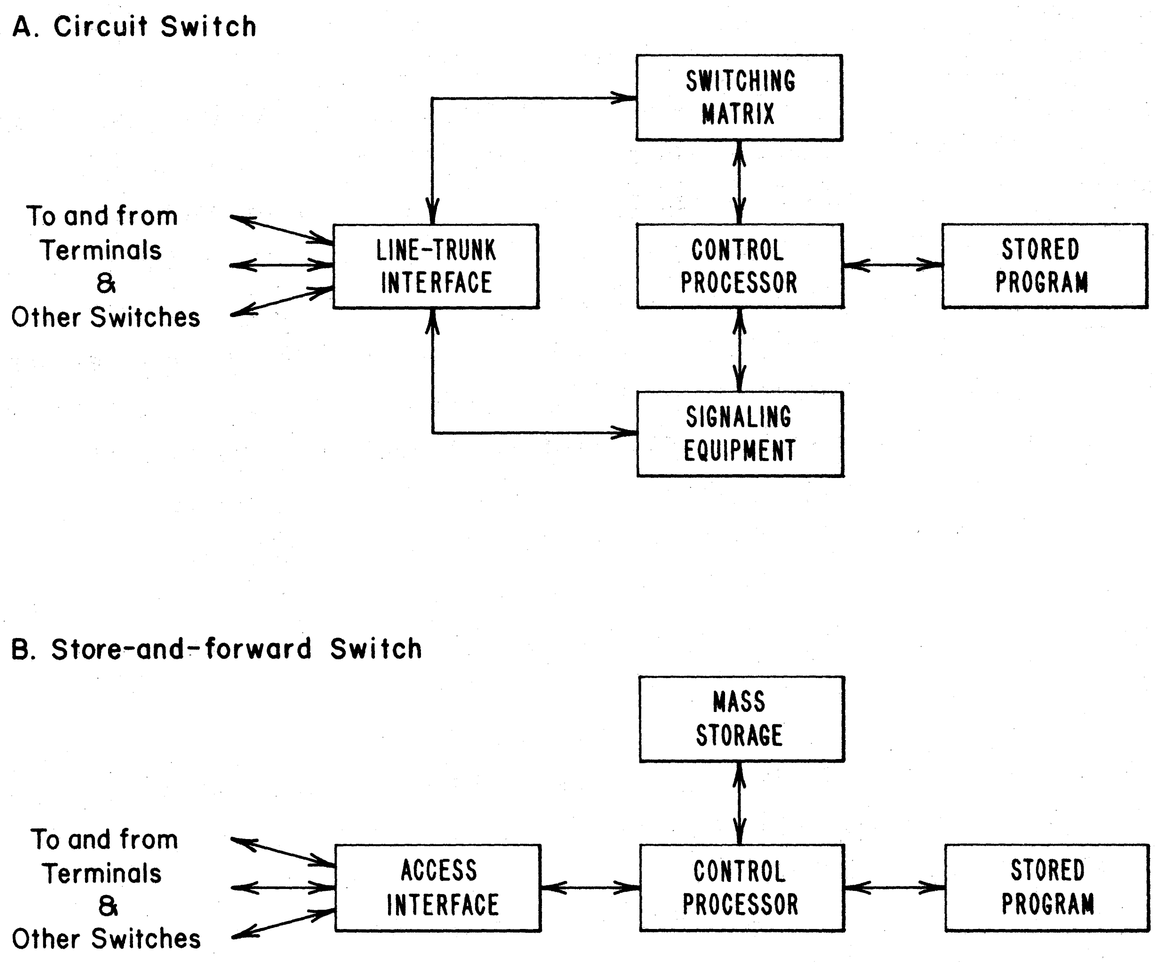


Figure 8. Block diagrams of two basic switch types.

stored program in the memory of a digital computer or processor. The circuit switch employs a time or space divided switching matrix whereas the store-and-forward switch employs mass storage facilities to hold information prior to forwarding. The control signaling information is included as overhead on each message or packet and is not shown as a separate element in Figure 8.

2.2. Interfacing and Signaling

In the previous section, it was shown how switches may be classified in terms of the technology used for the switching matrix and the control. In addition, the switch may be sized in terms of the number of ports or terminations it presents to the outside world. The interface between the switching matrix and the transmission facilities determine this size. In a digital circuit switch, the interface may be the principal element which determines the probability that a connection is blocked; i.e., the grade of service. These aspects of the interface are discussed in Section 4 and in Section 5.

In any given telecommunications network, the user terminals and the transmission facilities, as well as the switch, may be analog or digital. The interface at a given type of switch must provide the match to the transmission environment in which the switch operates. Figure 9 illustrates various combinations employed to interface a given analog or digital terminal to its particular local, either analog or digital switch, via analog or digital transmission facilities. All the paths between terminals and switches are considered bidirectional in Figure 9; i.e., terminal-to-switch follows same path back to terminal.

Caution -- to avoid misunderstanding, Figure 9 is to be interpreted literally, and in the following specific way only. The figure applies to a given unique terminal and its equally unique switch. Thus, if the terminal in question is analog and the end office switch is also analog, one uses the upper left part, (a), of Figure 9. There are two ways for this terminal to interface with its switch, namely, via analog or digital transmissions. Either one or the other, but not both, paths are permitted in the figure.

Tracing of loops is thus not to be done in Figure 9.

If the analog terminal is served by a digital switch, the same two transmission options exist, provided that CODEC's are located at the appropriate ends of the path.

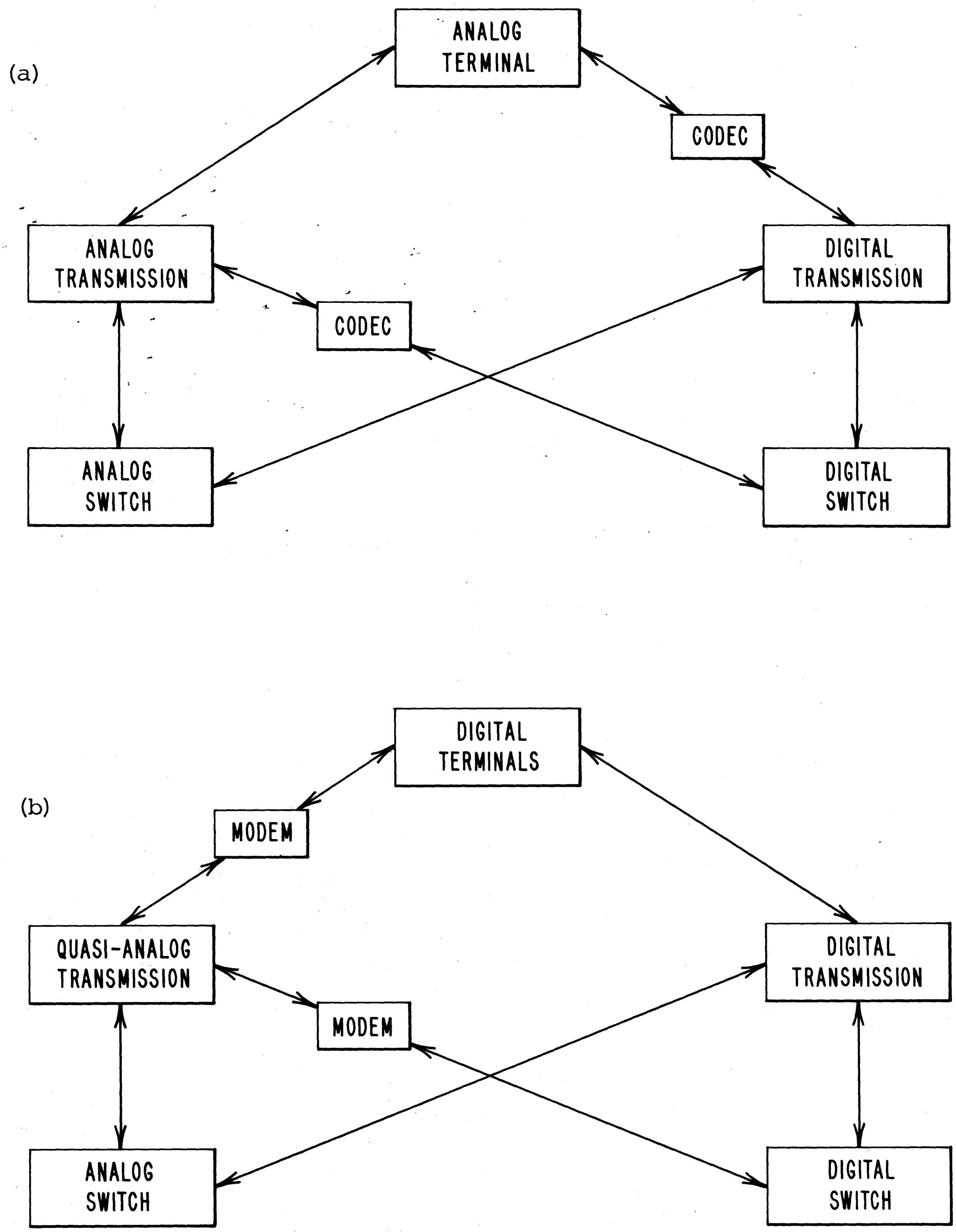


Figure 9. Interfacing digital and analog terminals with digital and analog switches.

If the terminal in question is digital, one turns to part (b) of the figure. Depending on the nature of the switch, either analog or digital, there are only two ways of placing the MODEM's to suit the two options for transmission facilities.

Analog signals must be A/D converted (coded) to be transmitted or switched in digital form. Likewise, they must be D/A converted (decoded) upon reception. This coding-decoding process is performed by the codecs shown in Figure 9(a).

Digital signals may modulate carriers to be transmitted or switched in analog (sometimes alluded to as quasi-analog) form and demodulated when received. This modulation-demodulation process is performed by the modems shown in Figure 9(b).

A given switch can operate in a dual environment using appropriate modems and codecs at the interface. A digital circuit switch with separate interfaces for digital and analog lines or trunks is described in Section 2.4.

The interface matches the transmission facilities to the switching element. It also provides the means for remoting the control element by means of the signaling element.

Signaling in telephony involves the exchange of electrical information (other than speech) which is specifically concerned with establishing, maintaining, and disengaging the connections. Signaling is also used to manage the network, to control the flow of traffic, monitor status and assess performance. Signaling systems provide the control information required to operate switches and terminals, ensure efficient and reliable transmission, and overall network administration. Table 1 lists several basic functions performed by the control signaling system.

The signaling required to remotely control a circuit switch is sometimes related to the interface. This occurs when user information and control information share the same transmission channel. When separate channels are dedicated for signaling, a separate interface is required. Signals may be transferred between terminals and switches (station signaling) and between switching points (interswitch signaling) using different technologies and modes of operation. Figure 10 provides a breakdown of the major signaling technologies used in circuit switched networks.

Table 1. Functions Performed by Various Types of Control Signals

Terminal Control Signals

- o Attending
- o Addressing
- o Alerting
- o Supervision

Switch Control Signals

- o Registration
- o Translation
- o Path Searching and Selection
- o Routing

Transmission Control Signals

- o Synchronizing
- o Error Detecting
- o Testing

Network Management

- o Flow Control
- o Status Monitoring
- o Performance Assessment
- o Reconfiguring
- o Recording

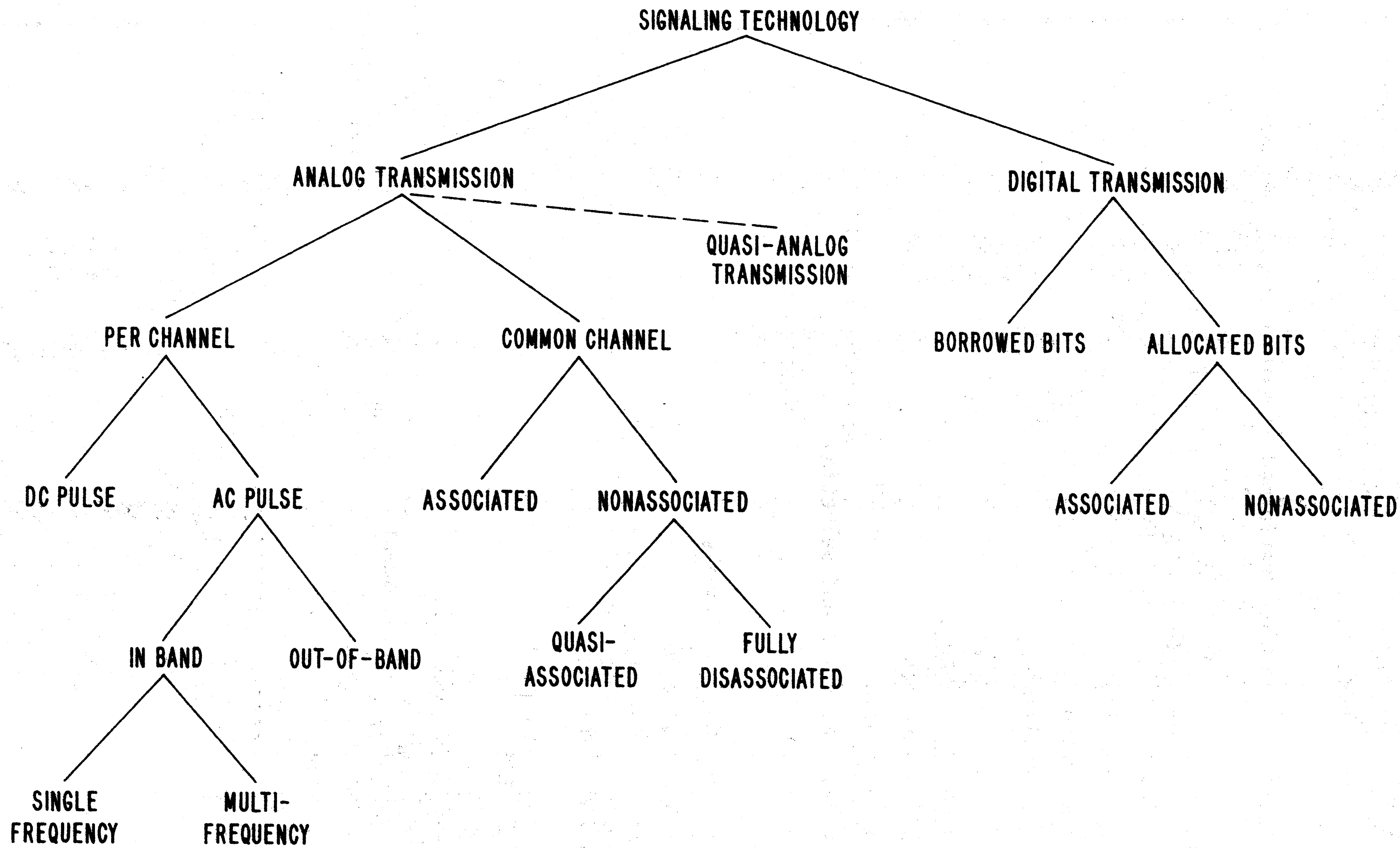


Figure 10. Signaling technologies used in circuit switched networks.

The control information inputs to the control element are, in most cases, generated by a process known as scanning. Outputs from the control element are dispersed by the distribution process. Distribution and scanning may be performed at the interface where signaling is accomplished on a per-channel basis. For common channel signaling and in store-and-forward switched networks, these functions are performed by the control processor itself.

Control signaling in a military switching environment is described in detail by Linfield (1979).

2.3. Switch Selection Criteria

The criteria for selecting the switch type for any given application is largely dependent on the nature of user terminals and their traffic demands. Terminals may have wide disparities in traffic rates, transaction sizes, delivery times and specific performance, such as synchronization, requirements. The transmission rates may vary from 100 bits per second for teletypewriters to megabits per second for wideband video and graphics. Transaction sizes range from less than 10 bits for character-oriented interactive codes to several megabits for bulk data transfers. Delivery time characteristics vary from continuous real time to intermittent data. Acceptable delays could be on the order of milliseconds, or minutes to hours for interruptable bulk data.

Different switch technologies may fit different kinds of traffic. The circuit switch is inherently a 2-way transparent switch and is ideal for handling continuous traffic with long holding times, such as on the order of one minute or longer. Circuit switches provide real-time connections between compatible terminals such as the telephone. Blocking due to congestion in the switch, or elsewhere, may occur.

The message switch is inherently one-way and is essentially non-blocking. Relatively long delivery delays may be encountered. Message switches are therefore suitable for handling terminals that generate interruptable types of traffic, such as facsimile or teletype. Terminals need not be compatible, since code, mode and speed conversion can be accomplished at the switch. The packet switch is well suited to handle interactive traffic that comes in short bursts. Such data flow is generated by the query and response interactions between host computers and their terminals.

Figure 11 summarizes the traffic types and the switch applications in terms of desirable network properties (Kleinrock, 1976a). The three sides of the triangle in this figure represent a network property which the user desires. These desirable properties are high reliability, large throughput and short delay.

At each apex of the triangle, any two desired properties can be achieved by neglecting (and in most cases, counteracting) the third property. Thus continuous real-time traffic in the lower right corner requires large throughput with low delay, but does not typically require high reliability. This is the situation for digital voice networks and can be handled by a circuit switch. Interruptable traffic, such as file transfer, appears at the peak of the triangle. This requires large throughput and high reliability, but can be delivered with long delay. The message switch is suited for handling this interruptable traffic.

It may, in the future, be feasible to combine the functions of all three types of switches into a hybrid switch capable of adapting to the different kinds of traffic. One such approach is described by Ross and his colleagues (1977), whereby circuit and packet switching are combined to integrate voice and data networks. A packetized virtual circuit for integrating voice and data on a packet switched network is also described by Forgie and Nemeth (1977). Usually, these integrated approaches sacrifice certain desirable properties for some classes of traffic. Gitman et al. (1977) discuss some of the issues involved in hybrid switching.

In the remainder of this report, emphasis is placed on digital circuit switches with stored program control. Functions performed by each basic element of the digital circuit switch are defined in the next subsections.

2.4. A Digital Circuit Switch and its Call Processing Functions

The basic elements of a digital circuit switch are shown by the block diagram in Figure 12. A detailed discussion of this switch configuration is given by Linfield and Nesenbergs (1978). The line/trunk interface provides the terminations for subscribers loops from station apparatus and for trunking to other switches. Analog terminations are converted to digital signals by

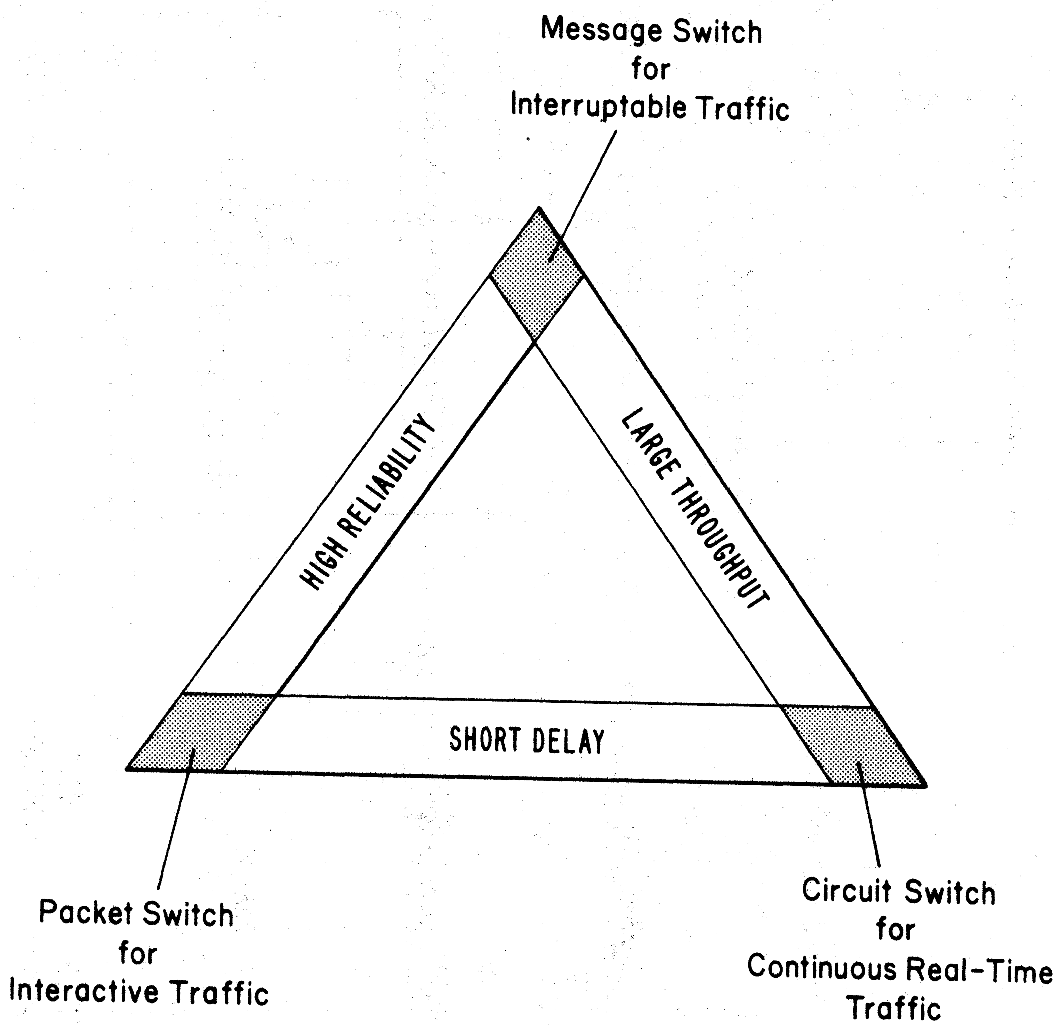


Figure 11. Switching system selection based on traffic classes and transmission properties.

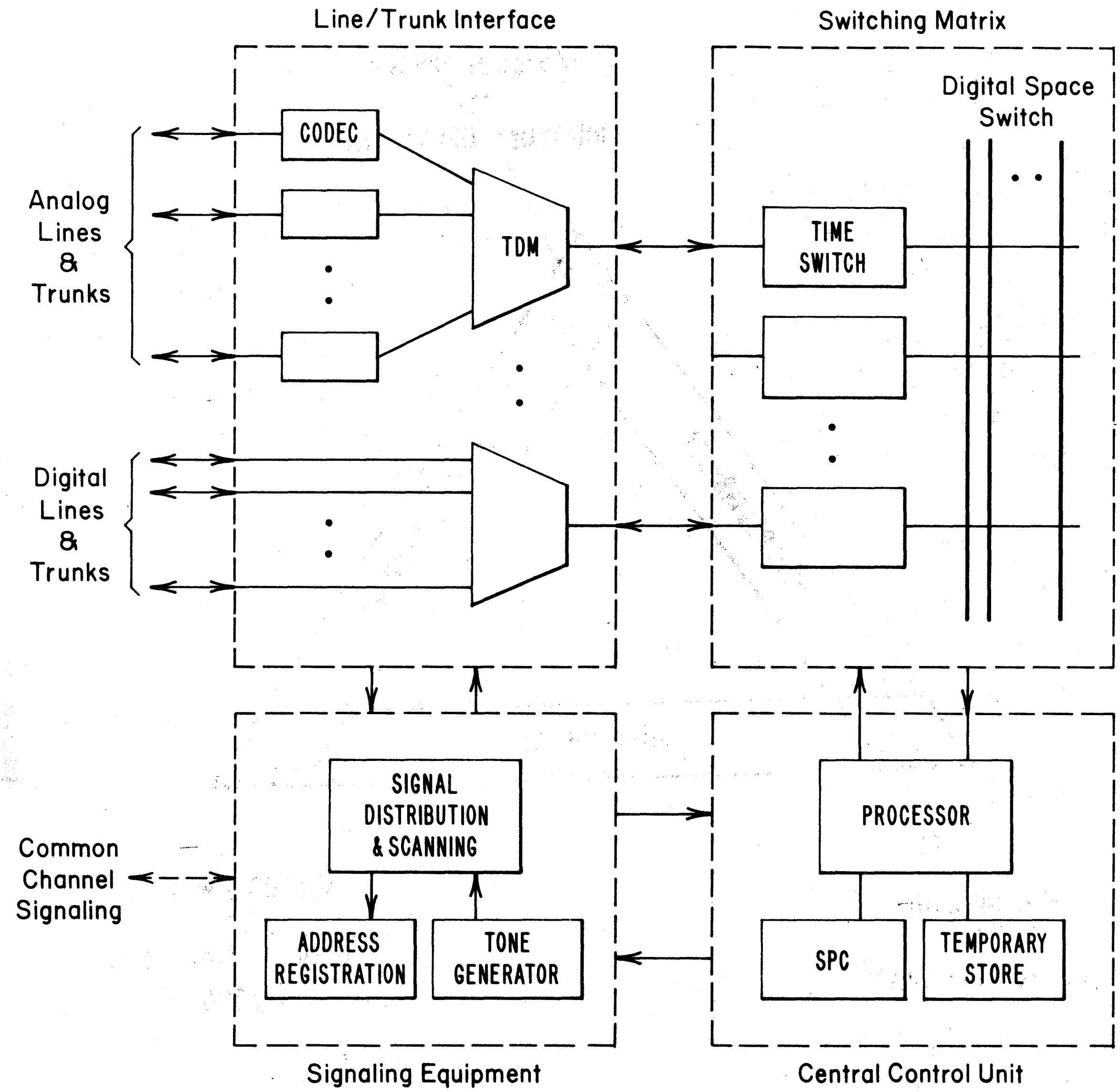


Figure 12. Digital switching system with T-S-T switching matrix.

A/D and D/A converters (codecs). Multiplexing or concentration may also occur at this point. The line interface also provides battery feed, overvoltage protection, ring access, supervision, clocking or codec functions, 2-wire to 4-wire hybrid connections, and test access, otherwise known as BORSCHT.

Interfaces for both analog and digital lines and trunks are shown in the figure. It is apparent from this figure that remoting the interface is quite feasible. Furthermore, the analog-to-digital conversion could be incorporated into the station apparatus with the multiplexer/ concentrator either remoted or colocated with the switch. Digital Secure Voice Terminals (DSVT) are feasible by adding encryption and decryption equipment to each terminal.

The interface also provides access for control signaling to the transmission facilities when signaling is provided on a per-channel basis. Such a signaling technique is assumed in the diagrammed system. Common channel interswitch signaling is also used. This would be indicated by a separate interface to a dedicated signaling channel. In Figure 12 the common signaling channel is shown as a dotted line.

Signaling equipment sends and receives information for controling the switch from remote terminals or from other switching points. This includes supervisory signals which indicate line and trunk status, address information, call progress tones, network management, and maintenance or administrative information.

The control unit interprets the signaling information, makes decisions, selects paths, and makes the connection via the switching matrix. The matrix shown in Figure 12 is a time-space-time matrix consisting of short memory devices to perform time slot interchanges and digital logic gates to interconnect the time-divided highways.

The call processing functions performed by major elements of the circuit switch are summarized in Table 2. For each element, the pertinent functions are divided into three categories as follows.

Per-line functions: required on a continuous basis.

Per-call functions: required only during the information transfer time.

Per-setup functions: required only during access and disengagement times.

Table 2. Call Processing Functions Performed by Circuit Switch Elements

	Signaling	Control Unit	Switching Matrix	Interface
Per-Line Functions (Continuous)	Attending (Service Request Detection)	Input Scanning		Battery Feed Overload Protection Hybrid
Per-Call Functions	Supervising (Call Completion Detection)		Interconnection	Codec Supervision (Answer and clear)
Per-Setup Functions	Addressing (Sending and Receiving) Alerting (Dial Tone, Ringing, Busy etc.)	Registration Translation Output Distribution Path Selecting Routing Busy Test Establish Connection Release Connection	Establish Connection Release Connection	Ring Access Test Access

Administrative and maintenance types of overhead functions which are not on-line traffic related, but which are performed by the processor when no calls are being processed, are not included in the table.

It is apparent that the control unit is required for call processing primarily during the initial call setup and subsequent disengagement conditions. One exception to this rule is the signal scanning process, which runs continuously, whenever per-line and per-trunk signaling equipment is used.

3. CHARACTERISTICS OF EXISTING CIRCUIT SWITCHES

Summary tabulations of the characteristics of some commercial type switches are given in Tables 3 and 4[1]. Specific values for many switches can be found in the literature or obtained from switch manufacturers. A review of several switching products is given by Pitroda (1977).

Toll, tandem and end office switch characteristics are summarized in Table 3. These exemplify only a few of the large range of sizes available in this category. More recently several switches have been developed which serve dual roles, e.g., toll and tandem or tandem and end office.

The PABX characteristics in Table 4 represent only a few of the kinds and sizes available today. Some switches are modular in design and one basic switch can be used to meet several sizing requirements.

The characteristics for two military type switches are tabulated in more detail in Tables 5 and 6. The TTC-38 is an example of tactical switching equipment being fielded today. The TTC-39 is a digital and analog circuit switch which, like the TTC-38, is a shelter-housed, tactical, automatic telephone control office. This switch is being developed for use in the early 1980's. Both switches are designed to meet the requirements for survivability, mobility, and other military environmental factors.

Some items in Tables 5 and 6 are left blank due to lack of information.

In subsequent sections, we indicate how the various switch parameters are related. Given certain switch characteristics others can be obtained from these relationships.

[1] The values given in these tables are representative of typical switches found in a given size range. The nominal values for capacity are composites of several switches of similar size.

Table 3. Typical Switch Characteristics for Toll, Tandem and End Offices

Switch Element \ Application	Very Large Toll	Large Tandem	Large End Office	Small End Office
<u>Interface</u>				
Terminations	100K T*	60K T	65K L**	1500 L
A/D	64 kb/s PCM	64 kb/s PCM	64 kb/s PCM	64 kb/s PCM
<u>Switching Matrix</u>				
Technology	Digital	Digital	Analog	Digital
Cross Points	Time and Space	Time and Space	Time and Space	Time
Carried Load (E)	50K	15K	5K	400
<u>Signaling</u>				
Lines	--	--	DC pulse, DTMF	DC pulse, DTMF
Trunks	Common channel (CC)	CC	MF or CC	MF or CC
<u>Control</u>				
Technology	SPC	SPC	SPC	SPC
Architecture	Centralized	Centralized	Centralized	Centralized
Capacity (BHCA)	500K	350K	100K	10K
Processor	Custom	Custom	Custom	Microprocessor
Word Length (Bits)	26	32	44	16
Storage (Bytes)	500K	250K		

*T = Trunks
**L = Lines

Table 4. PABX Switch Characteristics

Switch Element \ Switch Size	Very Large	Large	Medium	Small
<u>Interface</u>				
Terminations	7500 (L+T)	2400 L, 600 T	1000 L, 250 T	130 L, 30 T
A/D	64 kb/s PCM	64 kb/s PCM	64 kb/s PCM	64 kb/s PCM
<u>Switching Matrix</u>				
Technology	Digital	Digital	Digital	Digital
Cross Points	Time-Space	Time	Time-Space-Time	Time
Load (Erlangs)	5K	1500	250	45
<u>Signaling</u>				
Lines	DC Pulse, DTMF	DC Pulse, DTMF	DC Pulse, DTMF	DC Pulse, DTMF
Trunks	MF	MF	MF	MF
<u>Control</u>				
Technology	SPC	SPC	SPC	SPC
Architecture	Centralized	Decentralized	Decentralized	Decentralized
Capacity (BHCA)	25K	18K	9K	1.5K
Processor	Custom mini	Microprocessors	Microprocessor	Microprocessor
Word Length	17 bits	10 bits	8 bits	10 bits
Storage (Bytes)	200K	340K	500K	140K

Table 5. AN/TTC-38 Switching System Characteristics

General		
Model	:	AN/TTC-38
Introduced	:	1975
Application	:	Tandem/End office (Tactical)
Interface		
Min. Terminations	:	320
Max. Terminations	:	640
Load/Line (Erlangs)	:	
Adapter	:	2- or 4-wire
Multiplexer/Concentrator	:	
Signaling		
Lines	:	SF/DC supervision, DC pulse or DTMF addressing
Trunks	:	SF/DC supervision, DC pulse or DTMF addressing
Switch/Matrix		
Technology	:	Analog space division, 4 stage
Crosspoints	:	PNPN diodes (see note 1)
Carried Load (Erlangs)	:	180
Control		
Technology	:	SPC
Architecture	:	Central
Max. Capacity (BHCA)	:	2700 (600 lines)

Table 5 (cont.)

Processor(s)

Number and Type	GTE Custom (See note 2)
Number of Registers	
Word Length (Bits)	24
Instruction Length (Bits)	24+1 parity
Data Bus (Bits)	
Address Bus (Bits)	
Clock Rate (MHz)	12.5 MHz
Cycle Time (μs)	5.6
Access Time (μs)	5.6
Storage	Ferrite Core
o Semipermanent	65K (48K implemented)
o Permanent	
o Off-line	Paper tape
Instruction Set	
Call Processing Inst.	47K
Maint. & Adm. Inst.	
Peripherals	Remote page printer

Notes

1. Crosspoints 12 channel groups level FDM, 4-wire wideband (80 kHz channel) information.
2. A second processor/memory unit provides backup. Transfer interrupt period is approximately 100 ms.

Table 6. AN/TTC-39 Switching System Characteristics

General		
Model	:	AN/TTC-39
Introduced	:	1982 (tentative)
Application	:	Tandem/End Office
Interface		
Min. Terminations	:	150
Max. Terminations	:	660 (with specified mix of analog and digital switches)
Load/Line (Erlangs)	:	0.30
Digitizer	:	CVSD @16 or 32 kb/s
Multiplexer/Concentrator	:	
Signaling		
Lines	:	DC pulse or DTMF addressing plus others
Trunks	:	SF/DC supervision, MF addressing plus others
Switch/Matrix		
Technology	:	Digital and analog
Crosspoints	:	Time divided and space divided
Carried Load (Erlangs)	:	181 (600 lines)
Control		
Technology	:	SPC
Architecture	:	Centralized
Max. Capacity (BHCA)	:	3300 (600 lines)

Table 6 (cont.)

Processor(s)

Number and Type	: Litton L3050
Number of Registers	: Optional
Word Length (Bits)	: 32
Instruction Length (Bits)	: 32
Data Bus (Bits)	: 32
Address Bus (Bits)	: 32 (15 used)
Clock Rate (MHz)	: 16
Cycle Time (μs)	: 100
Access Time (μs)	: 2.5 (core memory)
Storage (words)	: Core 3M (typical), 131K (min.)
o Semipermanent	: Yes
o Permanent	: Yes
o Off-line	: Magnetic tape
Instruction Set	: 100
Call Processing Inst.	: 195K
Maint. & Adm. Inst.	: 30K
Peripherals	: TTY keyboard, CRT, mag. tape

Notes

Service features, special system functions, physical characteristics, cost, and performance data are not included in the tables. There are simply so many options available that listing them is prohibitive. Also, some of these characteristics either do not affect the system capacity or are only indirectly related. Those service features and system functions that do affect capacity are discussed in Section 4.

4. FACTORS LIMITING CIRCUIT SWITCH CAPACITY

No single parameter has been found adequate to define the total capacity of a switch. Individual capacities must be defined for individual switch elements. The maximum capacity of a given switch may well vary, depending on the environment in which it is operated.

Table 7 lists the major factors which affect the switch parameters. It is assumed for this table that the circuit switching matrix is controlled by a processor with a stored program. The matrix and the interface may both be analog or digital, and each be blocking or non-blocking. Station signaling (i.e., control signaling to and from a subscriber terminal) may be in-band or on a separate channel. Trunk signaling may be on a common channel or per-trunk basis.

The capacity of each switch element may be viewed in different ways depending on the element's function. Thus the termination capacity (number of lines and trunks) is interface and concentrator related. The call attempts capacity is control and signaling related. The load or traffic intensity capacity is concentrator and switching matrix related. Storage buffers and routing plans also affect capacity. Inadequacy of any element can limit the capacity of the total system. The ultimate capacity of the system is also related to performance. The technology and architecture used in an element, as well as between elements, may limit the speed of service due to intolerable delays or they may limit the grade of service due to blocking. A balance between time-shared equipment from a common pool and the per-line equipment requires knowledge about the traffic offered, traffic carried, and performance desired.

For example, the control processor must spend a certain amount of time on each call attempt. It also must perform certain administrative and main-

Table 7. Some Factors Limiting the Capacity of Circuit Switch Elements

	Interface Element		Switching Matrix Element	Common Control Element	Control Signaling Element
	Access Ports	Concentrator			
Parameter Defining Capacity	Terminations	Traffic Load Carried	Traffic Load Carried	Arrival Rate	Arrival Rate
Units	Number of Lines and Trunks	Erlangs	Erlangs	Attempts/Hour	Attempts/Hour
System Factors Factors Determining Maximum Capacity	Area Served Subscriber Density Calling Habits Features and Functions Offered	Traffic Offered Concentration Ratio Ratio of Local to Distant Traffic Number of Multidrop Lines	Architecture Availability Ratio of Line to Trunk Traffic Holding Time Constraints	Architecture Call Type Mix Program Structure Flow Controls Speed of Logic Memory Size Access Time Peaking Margin Service Features	Technique Format Operating Mode Signaling Rate Message Length
Performance Limiting Factors	Probability of Blocking Access Time Disconnect Time	Probability of Blocking	Probability of Blocking	Acceptable Average Delay (If Blocked)	Acceptable Average Delay (If Blocked) Error Rate

tenance functions which are not traffic related. As the number of connection attempts increases, the processor may become fully occupied. Additional calls then face connection delays that may or may not increase beyond acceptable limits. The processors' real-time capabilities thus limit the system capacity. This control capacity could be increased by increasing logic speed, lowering memory access time or by even reorganizing the software access structure.

Certain processing functions could be reassigned by changing the control architecture or by adding parallel equipments.

In the following subsections the switching element impacts on total system capacity are discussed.

4.1. Traffic Offered to the Interface

The proper design of a circuit switch requires knowledge of the expected traffic which the switch must handle. This traffic is characterized by the call arrival rate and the holding time statistics for the offered load, as well as the corresponding served call statistics for the actual carried load.

For a line that terminates at an interface, the incoming calls tend to occur randomly and vary in length. The arrival rate (λ) per unit time may be specified as the average number of busy hour call attempts (BHCA)[2]. The holding time (τ) is the average call duration.

These averaged parameters, λ and τ, determine the traffic intensity. Traffic intensity or offered load is the amount of traffic in a path or group of paths per unit time. It is higher than the actual carried load. The traffic carried over a single subscribers line is indicated in Figure 13. For long averaging periods, T, the unexpired calls occurring at the beginning and end of the averaging time have a negligible effect. Then the following definitions apply.

$$\text{Average holding time, } \tau = \frac{t_1+t_2+t_3+t_4+\ldots t_n}{n}$$

---

[2]The BHCA is often determined by averaging busy hours of traffic data from 20 days during a busy season. Day-to-day and hourly variations may be accounted for by adding a safety margin to the systems' capacity.

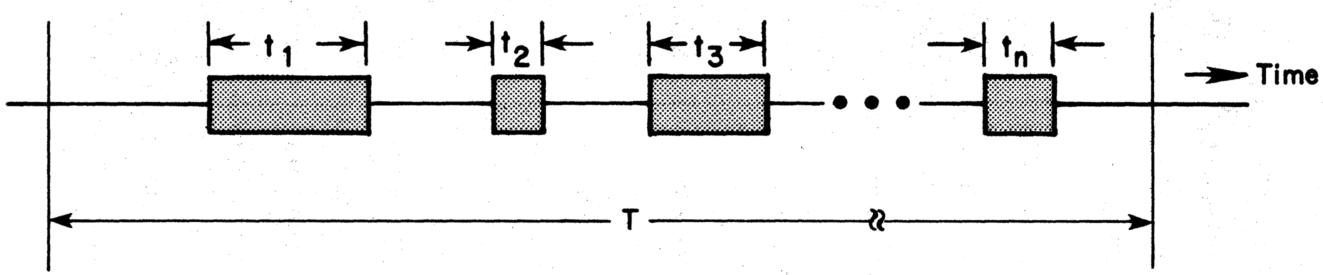


Figure 13. Line occupancy for a number of calls.

Average attempt per unit time $\lambda = n/T$

Average traffic intensity $A = \lambda\tau = \frac{t_1+t_2+t_3+\ldots t_n}{T}$.

These are for the load carried by the lines and offered to the switch. The international unit of traffic intensity[3] is the Erlang (E), a dimensionless quantity. One Erlang of traffic intensity on one line implies a continuous occupancy of that line. Therefore,

$$1E = {}^{1\text{ call hour}}/_{\text{hour}} = {}^{1\text{ call sec}}/_{\text{sec}}.$$

In North America, traffic intensity is usually expressed in terms of hundred call seconds per hour and abbreviated as ccs, with the per hour implied and not stated. Erlangs and ccs (per hour) are related as follows

$$1E = 36\text{ ccs}.$$

The traffic intensity offered per line as a function of call attempts per hour and holding time per call is illustrated by the lines in Figure 14. The maximum intensity carried by any line cannot exceed one Erlang.

The actual traffic intensity is a function of the subscribers calling habits. These habits may range from a few calls per day to several per hour. Holding times may range from less than a minute to an hour or, in rare instances, more than an hour. Variations in subscriber calling habits also occur. These variations depend on time of day, day of the week, season of the year, subscriber location (business versus residential), and other factors. In summary, the traffic intensity may vary from 0.01 Erlangs/line for residential subscribers to greater than 0.5 Erlangs/ line for some business subscribers. Trunk traffic may approach full occupancy at times.

The average busy hour load for a subscriber's line is typically about 0.1 E or 3.6 ccs. From Figure 14 this 0.1 E is found to correspond to one 6-minute call per hour, two 3-minute calls per hour, or six 1-minute calls per hour.

---

[3]The terms "traffic intensity" and "traffic load" are used interchangeably in this report and mean the quantity of traffic in one or more paths per unit of time.

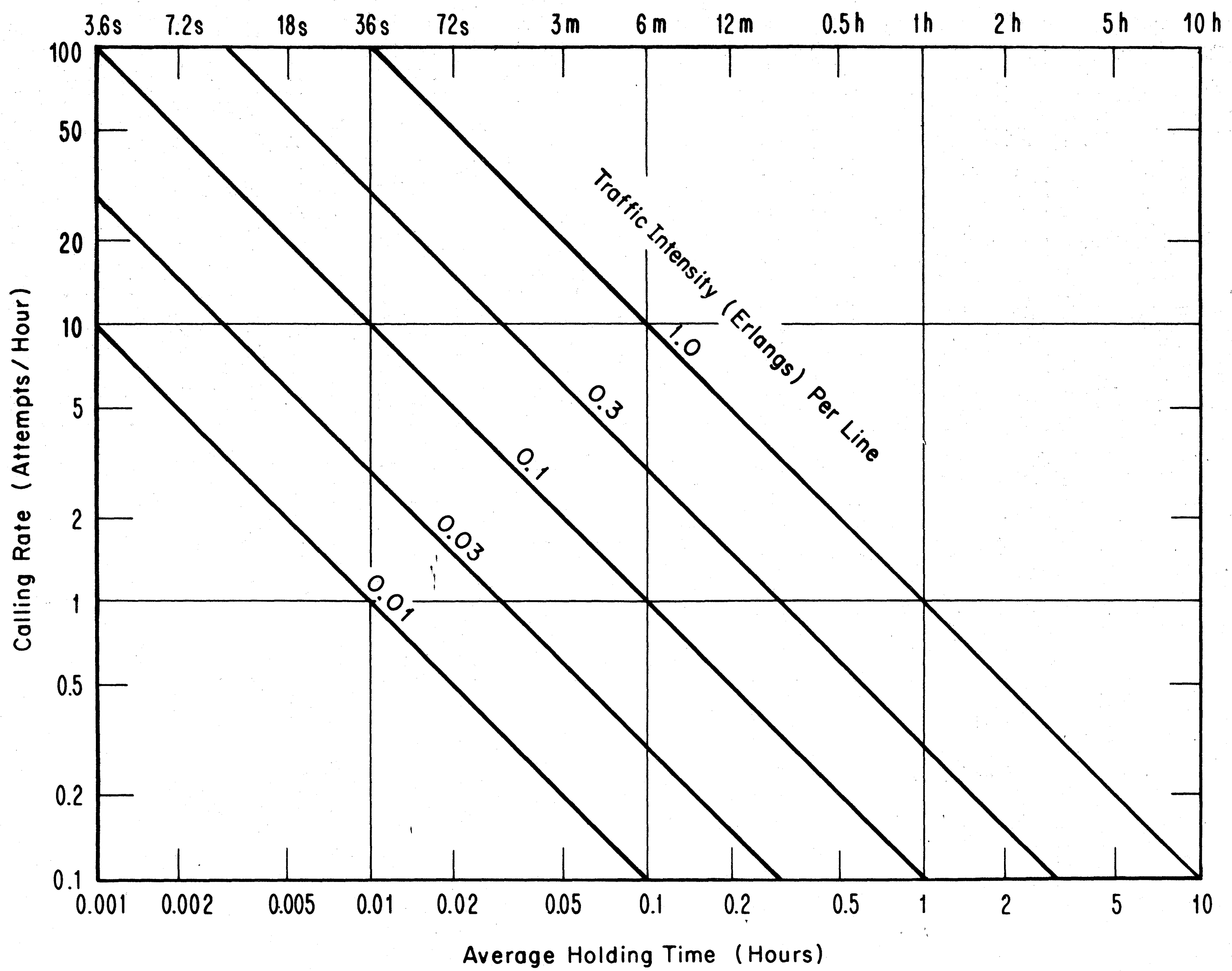


Figure 14. Traffic intensity per line as a function of call rate and holding time.

The shaded areas in Figure 15 indicate the approximate operating ranges for line and trunk traffic for a typical circuit switched network. There are, of course, overlaps between these two kinds of traffic.

Interactive traffic and bulk traffic operating ranges are also indicated on this figure. Again there may be considerable overlap. The interactive traffic occurs between data terminals that send and receive many messages with very short durations. Bulk traffic includes file transfers. Here connections are held for long periods of time, perhaps on the order of hours. In both of these cases (interactive and bulk), the use of a circuit switched network is uneconomical if they occur on a permanent daily basis. It is often more practical to either lease dedicated circuits from a carrier or to use different switching techniques.

4.2. Interface Multiplexing and Concentrating

One of the potential interface functions is multiplexing or concentration. It was noted previously that the maximum traffic intensity offered per subscriber's line cannot exceed one Erlang. Several lines can be multiplexed together onto a common transmission or trunking facility from a remote cluster of terminals, or at the switch for switching purposes. Such a common facility cannot carry a load in excess of 1 Erlang/trunk, although its offered load need not be limited.

The concentration entails blocking. However, the traffic intensity is usually much less than one Erlang per terminal. Often it is less than 0.2 Erlangs. Since this traffic is generated randomly, and full availability is desired, it is practical to use one or more concentration stages ahead of the switching matrix. This permits a variety of arrangements of additional lines to be switched at given times. The network is designed to provide an acceptable blocking probability to the user. The concentration ratio used is on the order of 2:1 to 6:1, depending on the number of lines in the group, the traffic intensity per line, and the grade of service desired. Also, all of the terminations to the switch may not be subscriber lines. Usually some calls are connected to other switches via trunks. The ratio of intraswitch calls to interswitch calls depends on the community of interest served by the switch. Concentrators may be remoted to areas of high terminal density. The remote concentration units (RCU) must be designed with an acceptable blocking

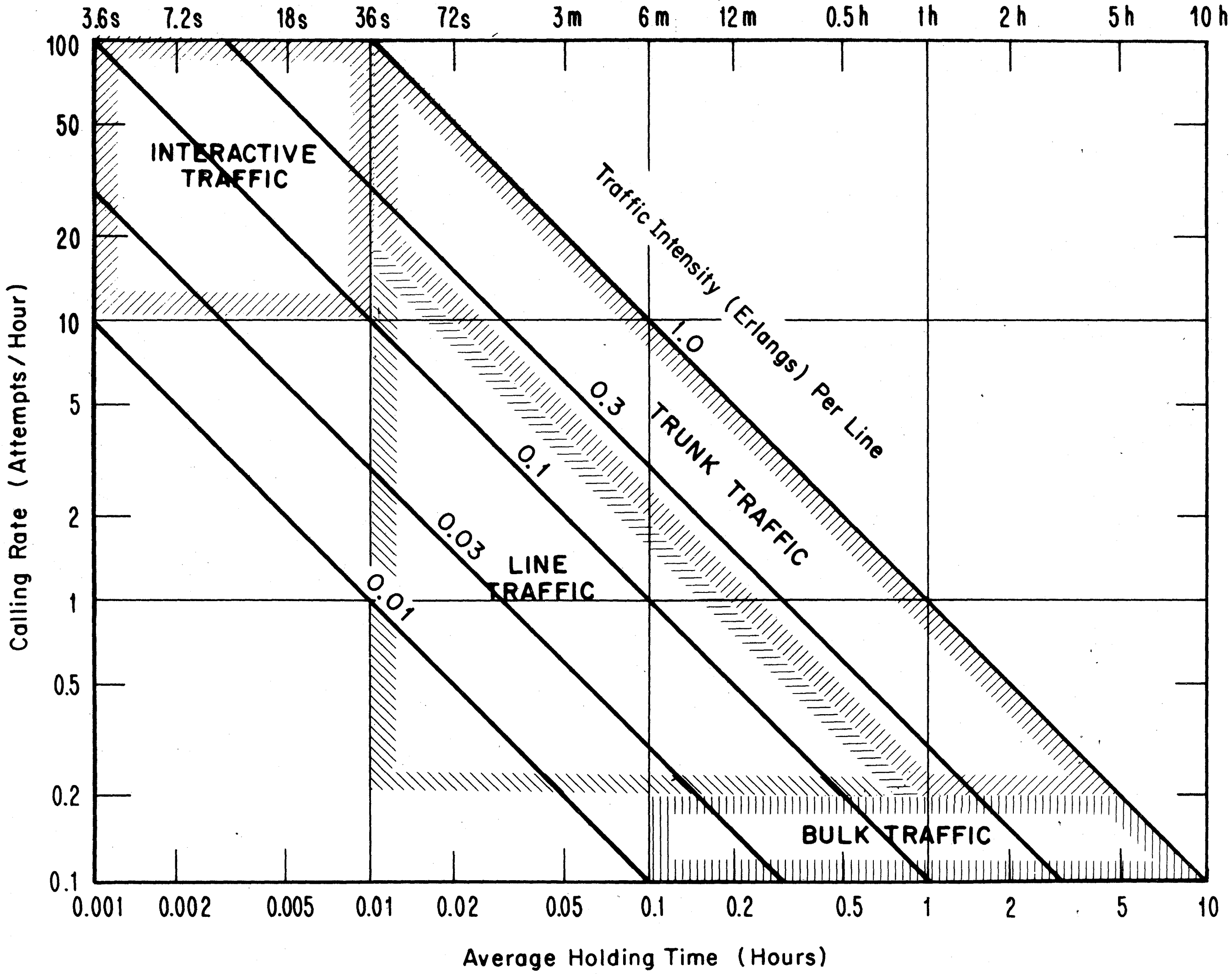


Figure 15. Approximate operating ranges for line and trunk traffic.

probability. Their performance still is a function of the community of interest in which they serve (see Sec. 5).

When traffic from a number of terminations is multiplexed together, the traffic intensity in Erlangs on the multiplexed output equals the sum from all input terminations. This is because the ordinary multiplexer allocates a fixed channel to each and every termination. This allocation may be by wire pairs in a cable (space division multiplexing), by frequencies in a channel (frequency division multiplexing), or by time slots on a highway (time division multiplexing). If the average traffic intensity per line is A, then the multiplexed intensity is MA, where M is the number of input terminations.

When traffic from a number of terminations is concentrated, the maximum traffic intensity at the concentrator output is less than or equal to the number of input channels. Input channels are allocated dynamically by the concentrator. Fewer output channels are required, since they are assigned as needed. When all output channels are occupied, new arrivals are blocked.

For a fixed probability of blocking, the output capacity in Erlangs decreases as the concentration (ratio of input to output channels) increases. The effect due to different number of sources is indicated in Figure 16. The curve in the figure shows the traffic intensity or capacity of 10 output trunks in Erlangs, as a function of the number of input terminations. It is assumed that on the average 1 in 100 distant calls is blocked and cleared[4]. It is seen, for example, that 50 lines may offer 4.5 Erlangs of traffic to the 10 channels with a P_D=0.01 blocking probability. The average traffic intensity per input line must therefore not exceed 4.5/50=0.09 Erlang for the specified grade of service.

In a digital switching system, the line/trunk interface may include a coder and decoder (codec) for analog inputs, and a time-division multiplexer or concentrator, as shown previously in Figure 12. The output bus from the concentrator or multiplexer may contain any number of time slots.

First level multiplexers, often used in North America, digitize speech with a pulse code modulation (PCM) developed for T-carrier transmission facilities. The codec samples the speech signal 8000 times each second and encodes each sample with an 8-bit binary number. The resulting bit rate for

[4]Distant calls and distant call blocking probability, P_D, are defined in Section 5.2 and in Appendix A.

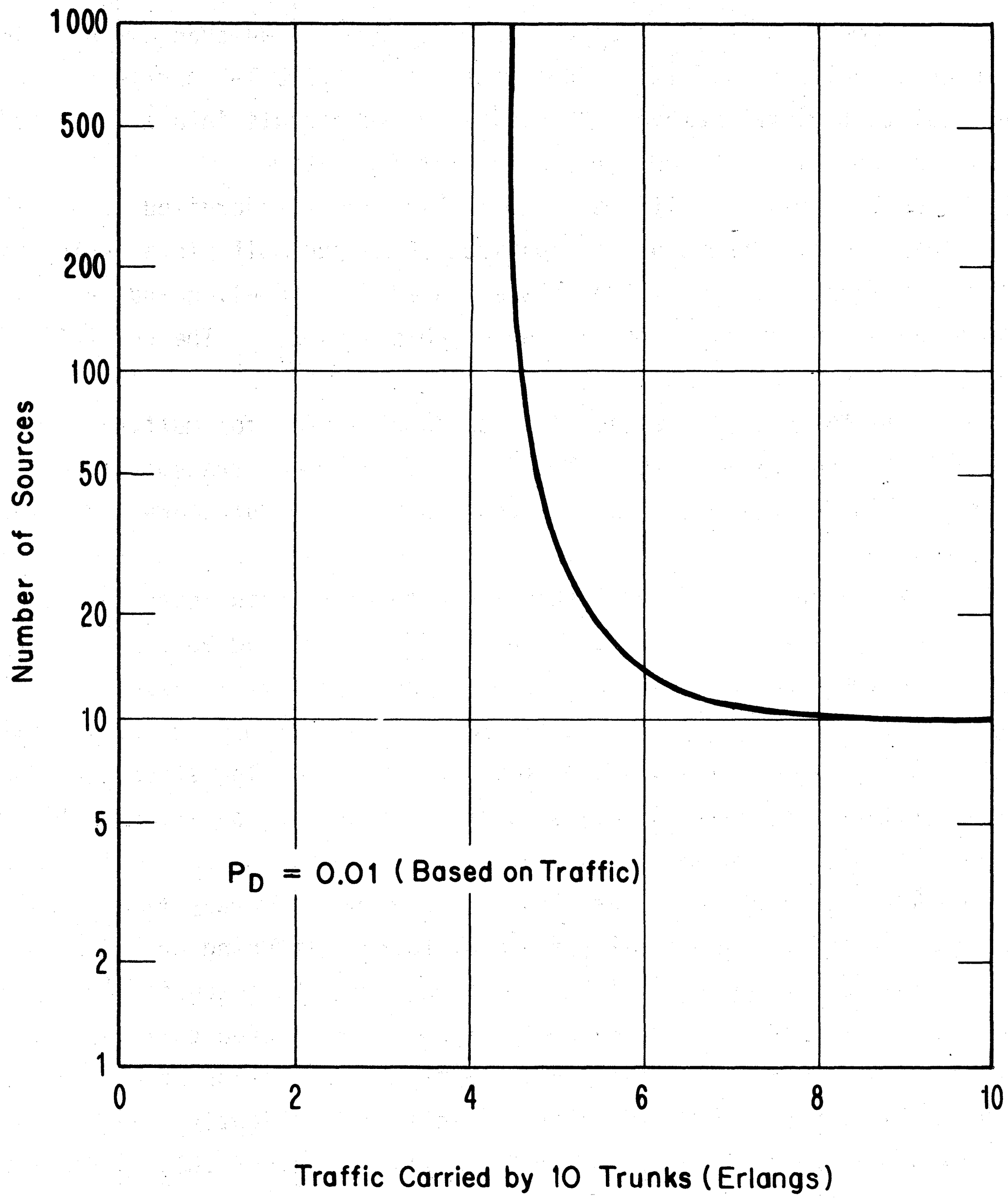


Figure 16. The effect of the number of sources on capacity of 10 trunks with 0.01 blocking probability and assuming blocked calls are cleared.

a speech channel is therefore 64 kb/s. Twenty-four channels are assembled into a frame of 192 bits. One bit is added for frame synchronization, resulting in a transmission rate of 8000x193=1.544 Mb/s. Control signaling information is transmitted by borrowing the least significant bit from each channel every sixth frame. The digital signal output from this 24-channel multiplexer is known as a DS-1 signal. It can be transmitted over a T-1 carrier. A second level multiplexer combines 96 digital voice signals into a 6.312 Mb/s, DS-2, signal, which, is suited for T-2 carrier facilities.

In Europe, a somewhat different scheme has been standardized (see CCITT, 1977a). Thirty voice channels are digitized (PCM) and multiplexed using time division techniques. Two more channels are used for signaling and synchronizing, bringing the total to thirty-two 64 kb/s channels. The resulting transmission rate is 2.048 Mb/s.

Similar techniques can be used for concentration as for multiplexing. The concentrator output may use either T-1, T-2, or other transmission facilities. Digital repeaters may be required when concentrators are remoted from the switch.

The curves drawn in Figure 17 show the offered traffic intensity per line as a function of concentration rates for 4, 10, 24 and 96 channel outputs. The later two correspond to T-1 and T-2 type carriers. It is assumed that blocked calls are cleared and the concentrator has a one-way blocking probability of 0.01. It is also assumed that each concentration stage has a low community of interest; i.e., inputs are seldom, if ever, connected to other inputs on the same concentrator.

A concentrator with 96 two-way input lines and 24 two-way time-division output channels has a concentration ratio of four. Referring to Figure 17, it can be seen for this example that the average traffic intensity per line cannot exceed 0.2 Erlangs, or more than one out of a hundred call attempts will be blocked. For connections involving two concentration stages, as is often the case, the two-stage blocking probability is slightly less than the sum of the two probabilities. If each has a probability of .01, the two-stage blocking probability is roughly .02.

When a concentration stage serves subscribers with a high common community of interest, the blocking probability is different than that indicated in Figure 17. This situation is discussed in Section 5.

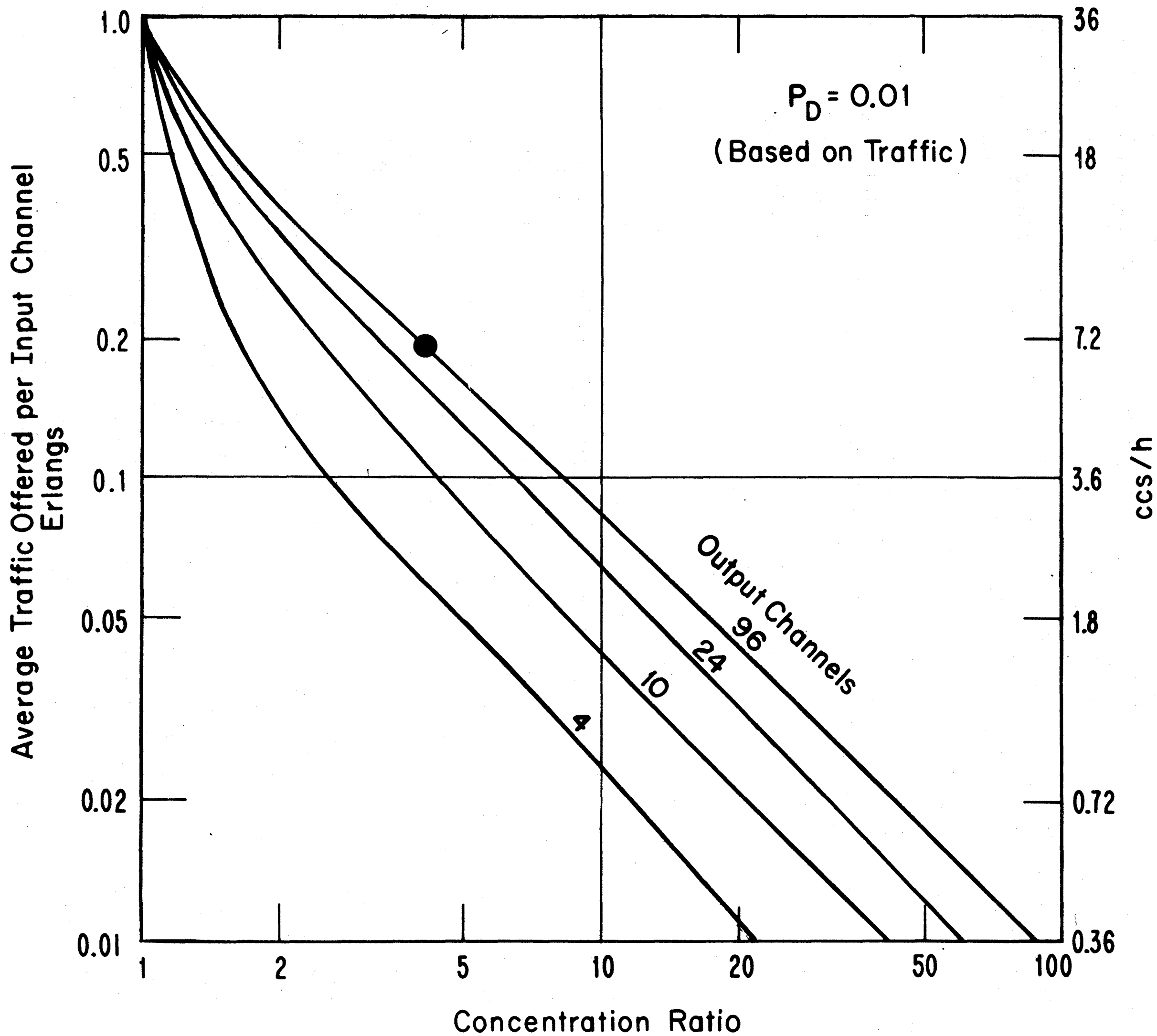


Figure 17. Concentration achievable for one-way blocking probability of 0.01 and assuming blocked calls are cleared.

4.3. Traffic Circulation in the Switch Matrix

The digital switch depicted by the block diagram in Figure 12 incorporates a T-S-T switching matrix. Any combination of lines and trunks can be terminated at the interface and interconnected by the matrix. The matrix is assumed to have full availability to all access channels. The purpose of this section is to define the traffic handling capabilities or capacity of the matrix.

The total traffic entering and leaving the matrix during any given period must be equal. Therefore the total traffic carried by the switch is

$$A_T = \frac{mA}{2}$$

where m is the total number of access ports on the matrix and A the average traffic intensity per access port. For a 300 line switch and assuming 0.15 Erlangs/line, the maximum traffic carried is the offered load

$$\frac{300x0.15}{2} = 22.5 \text{ Erlangs.}$$

This assumes that all traffic is intraswitch traffic and none originates or terminates on another switch via trunks.

When trunk traffic is included, the carried traffic may be differently estimated. It depends on how the terminations are counted. For example, assume that on the average each subscriber originates 5 calls per hour. Furthermore, assume that 2 of the 5 calls are to other subscribers homed on the same matrix (intraswitch calls), and 3 are to other subscribers homed on a different switch (interswitch calls). During the hour, the same average subscriber may receive 4 calls, 2 of which are intraswitch calls and 2 are interswitch calls. Let the average holding time be one minute. Then the average traffic intensity on that subscriber's line is as follows.

Originating Intraswitch calls	:	0.033 Erlangs
Originating Interswitch calls	:	0.050 Erlangs
Received Intraswitch calls	:	0.033 Erlangs
Received Interswitch calls	:	0.033 Erlangs
Total Intensity per line	:	0.150 Erlangs

The traffic circulation in the switch matrix for this average subscriber is shown in Figure 18.

If only the lines are counted in sizing the switch, then for a 300-line matrix the total traffic carried is, as before, 22.5 Erlangs. When the trunk

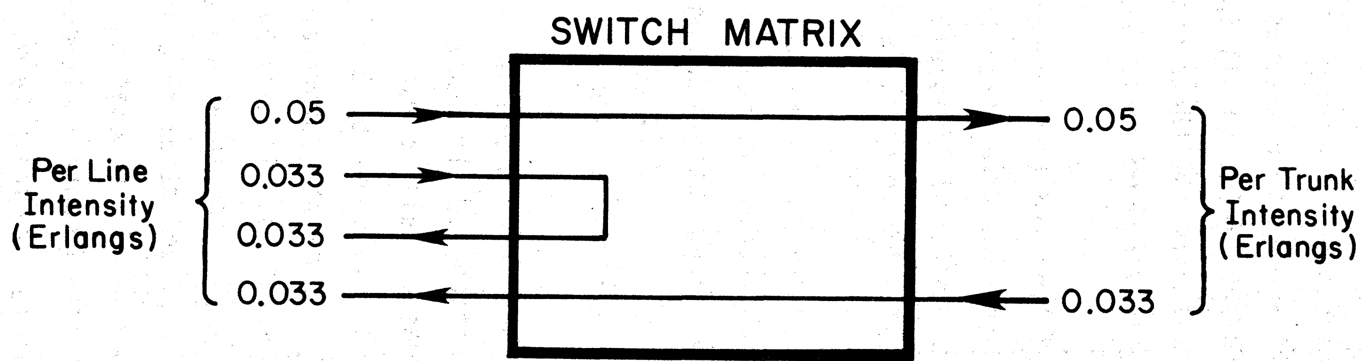


Figure 18. Average traffic circulation per line and per trunk for a specific example.

calls are included in the fixed line load, the carried traffic increases. Thus it is seen, from Figure 18, that the actual traffic carried by the switch is

$$300(0.05+0.033+0.033) = 35 \text{ Erlangs.}$$

If all of the line's 0.15 Erlangs were through traffic, then the total traffic carried by the switch would be 45 Erlangs.

Figure 19a summarizes this example for all intraswitch, all interswitch, and combined traffic conditions. In all three cases, the switching matrix has 300 input lines and is considered non-blocking. Trunks are added as required.

When the total number of line and trunk terminations is constant, the non-blocking switch matrix carries a fixed traffic intensity. This is shown in Figure 20, where the total number of lines plus trunks is held constant at 300. The traffic intensity per line or trunk is 0.15 so the matrix capacity is 22.5 Erlangs.

In actual practice, the number of trunks may be reduced considerably by accepting some finite blocking probability for interswitch calls. For example, in Figure 19, the 15 Erlangs outgoing and 10 Erlangs of incoming trunk traffic could be carried by 28 and 21 trunks, respectively, assuming blocked calls cleared, full availability, and a trunk blocking probability of 0.001.

A set of straight lines have been drawn in Figure 21 to relate the call attempts per hour, per termination, to the number of originating terminations, for different parametric values of calling rates or number of call attempts per hour. The same relationship is shown by the curves in Figure 22, except the calls per termination is used as the parameter. There is a reason for using both parametric representations. It will be clarified shortly. One additional relationship is needed: the total number of call attempts per hour versus average holding time, and parametric in matrix capacity. This relationship is shown in Figure 23. The total number of call attempts handled by the switch depends on the common control element; i.e., on the processor's capability. This is indicated by the ordinate of Figure 23. Constant processor capacities appear as horizontal lines across this figure.

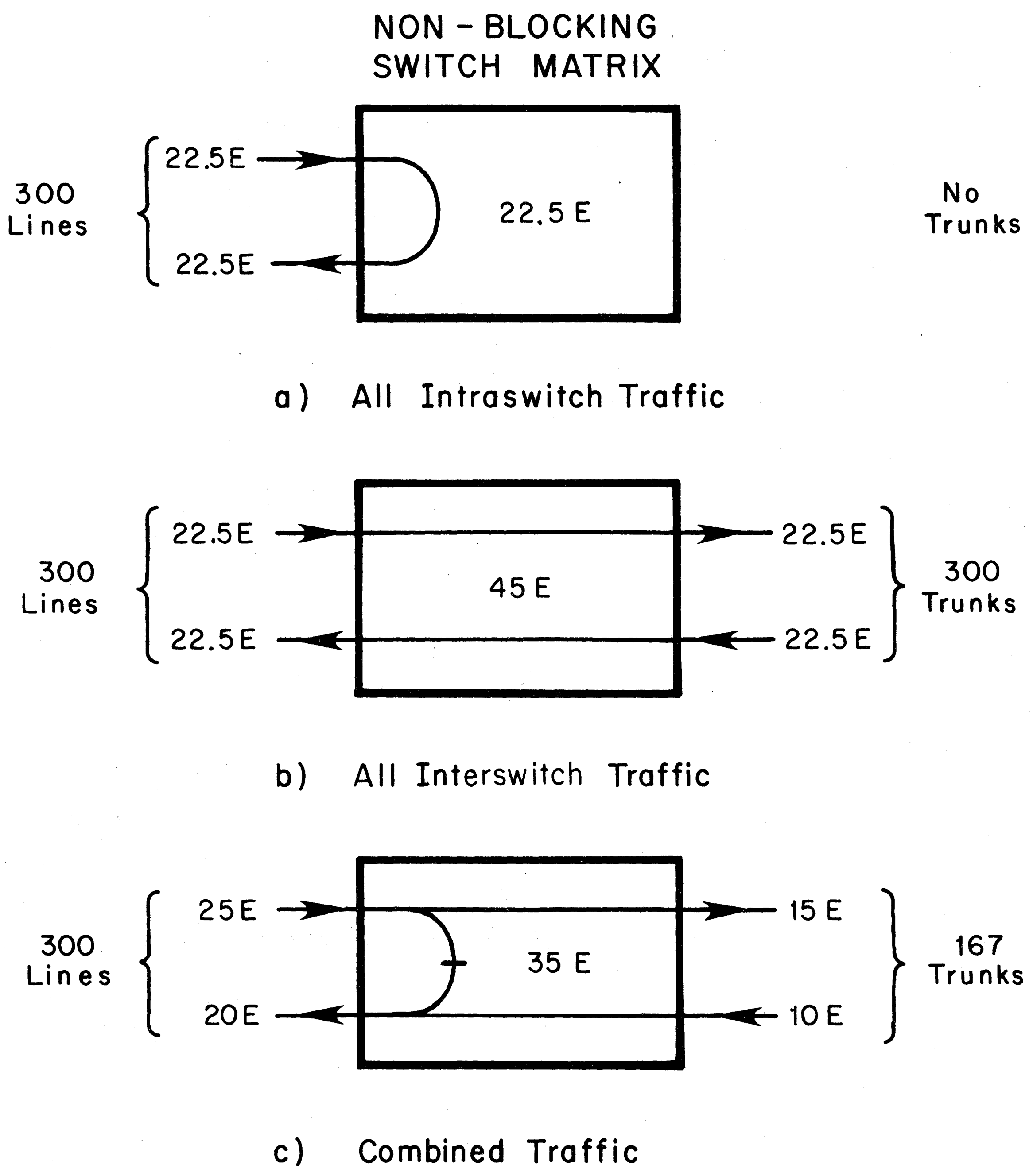


Figure 19. Traffic circulation in switching matrix as a function of community of interest. Line ports fixed at 300 and trunk ports added for interswitch traffic.

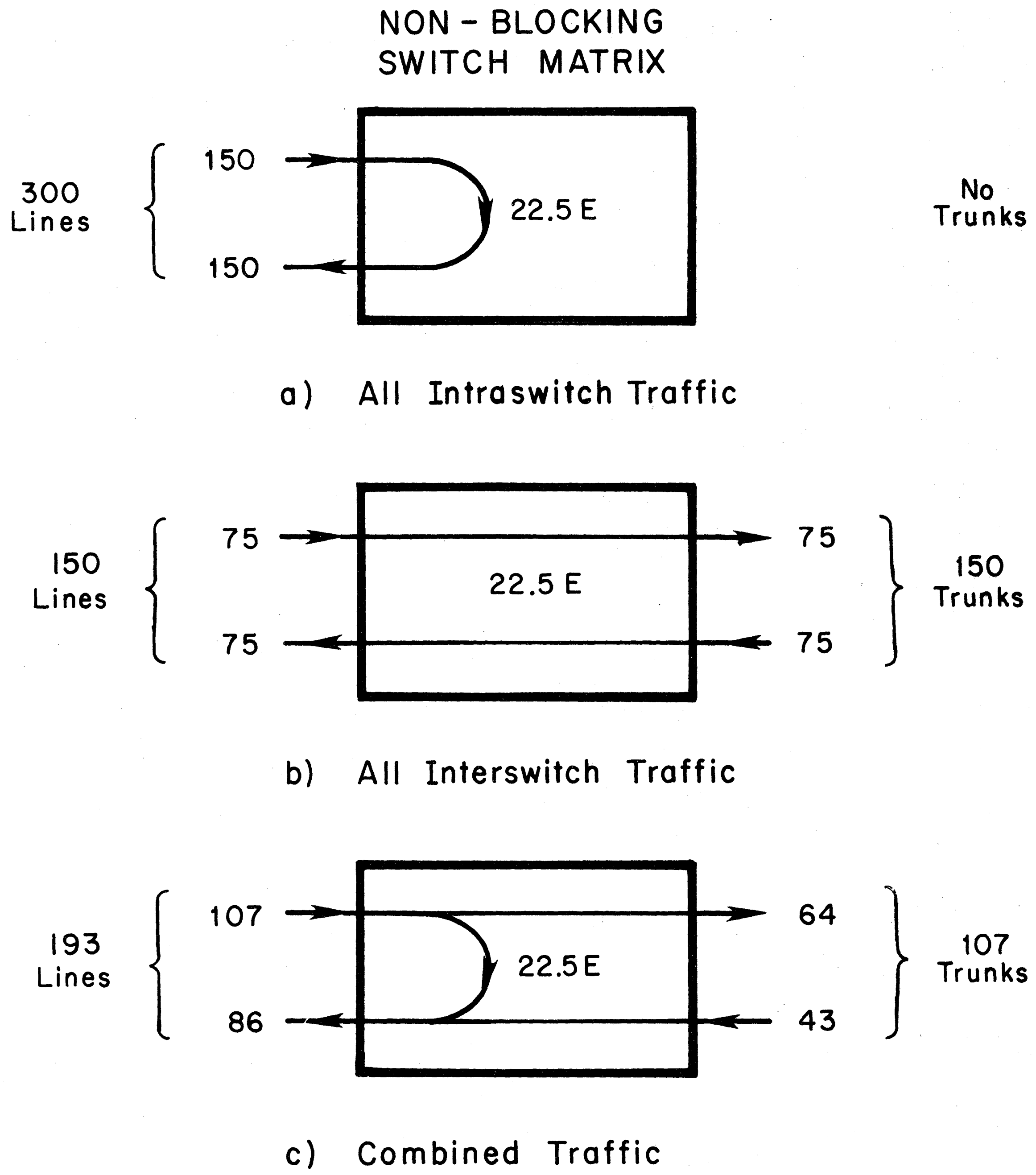


Figure 20. Traffic circulation in non-blocking matrix with total number of lines and trunks fixed at 300.

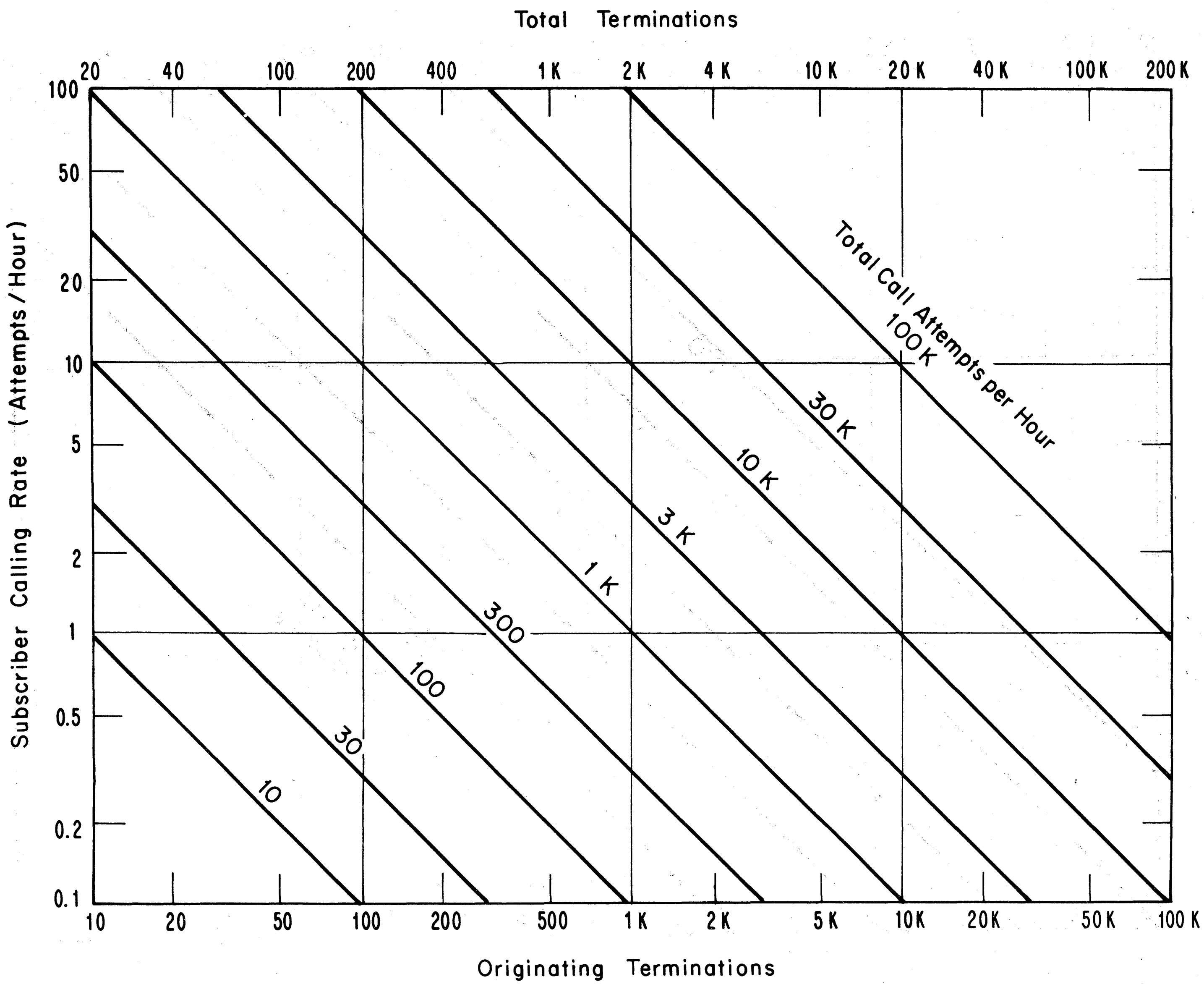


Figure 21. Relating average subscribers call rate to total call attempts.

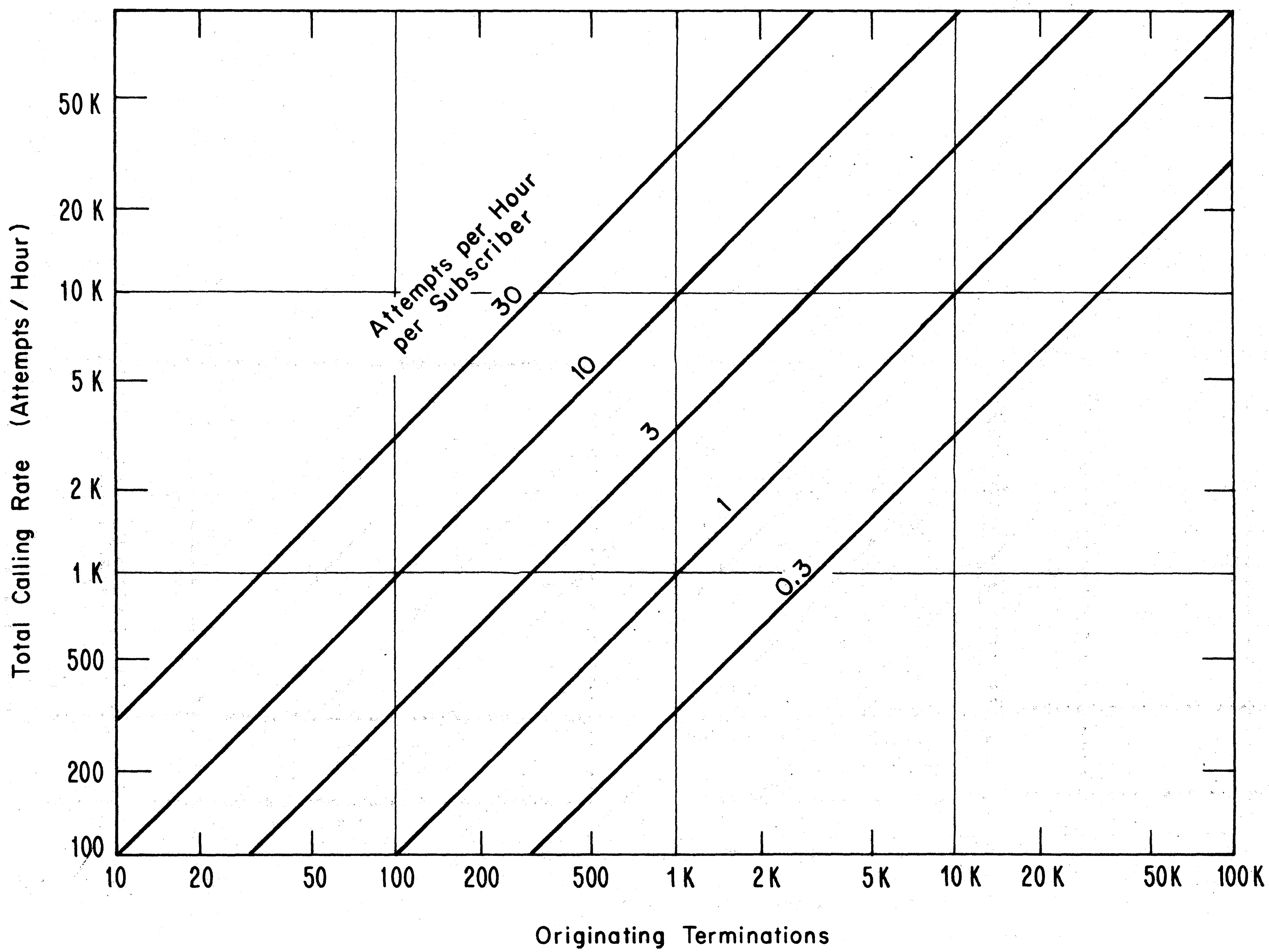


Figure 22. Relating total call attempts to number of terminations.

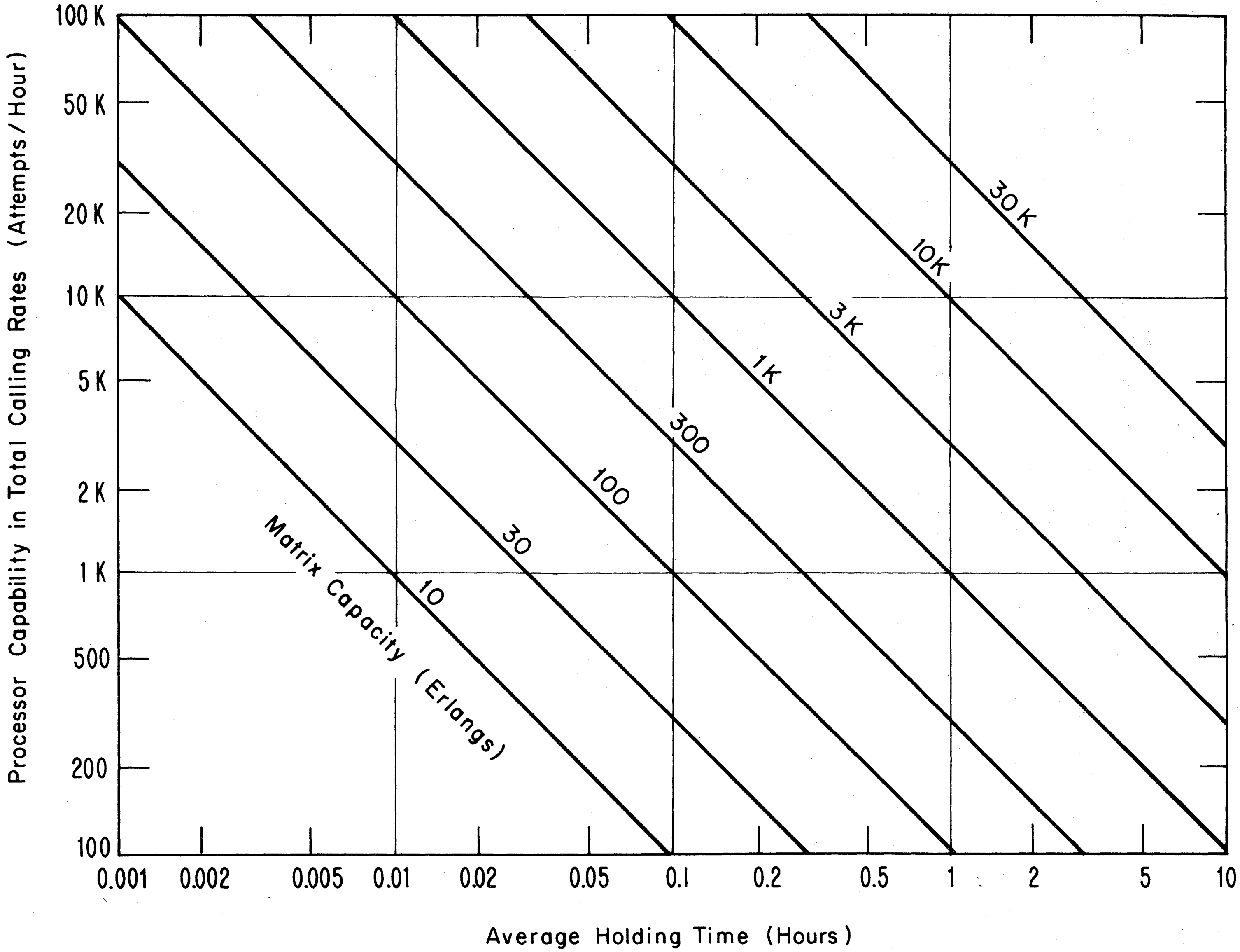


Figure 23. Relating processor capabilities to matrix capacity.

The interrelationships between calls per line, number of lines (and trunks), switch matrix load carrying capacity, and processor capacity are developed in the next section.

4.4. Interrelationships Between Capacities of Switch Elements

Figures 14, 21, 22 and 23 can be combined into a single figure by matching the common abscissa and ordinate scales. This composite, as shown in Figure 24, is useful in evaluating various switch configurations. Also, since different manufacturers specify switches in different ways, the curves in Figure 24 can be used to convert different specifications to a common base. Figure 24 also aids in sizing a switching system and in determining the relationship between processor capabilities, traffic intensity, matrix capacity, and the number of terminations. These applications of Figure 24 are demonstrated by the examples given in the following paragraphs.

Some of the switch characteristics given in Tables 3 through 15 were derived with the aid of Figure 24. As example of this procedure, consider the SL-1 digital switch (Telesis, 1976).

The SL-1 switch is a PABX. Modular sizing permits applications in the range from a few hundred to several thousand switch terminations. The maximum number of terminations is given as 7600. This size is considered here.

The SL-1 concentrates input terminations onto time division multiplexed loops. The interface concentrator digitizes voice channels using standard PCM at 64 kb/s per channel. Each loop has 32 time slots, 30 for traffic, one for control signaling, and one for spare. The loop data rate is therefore a CCITT standard but a USA non-standard, 32x64=2.048 Mb/s. A time slot interchange permits switching between time slots within the same loop. Multiple loops and a space division network provides large numbers of time slots and terminations. Time-multiplexed switching permits interchanging channels between the loops. Large switches combine 16 loops into groups, 12 for traffic and 4 for service circuits. Very large switches may consist of up to 5 groups. They need additional memory and multiplexers to perform switching between groups. The maximum size switch therefore contains 30x12x5=1800 traffic channels at the concentrator's output.

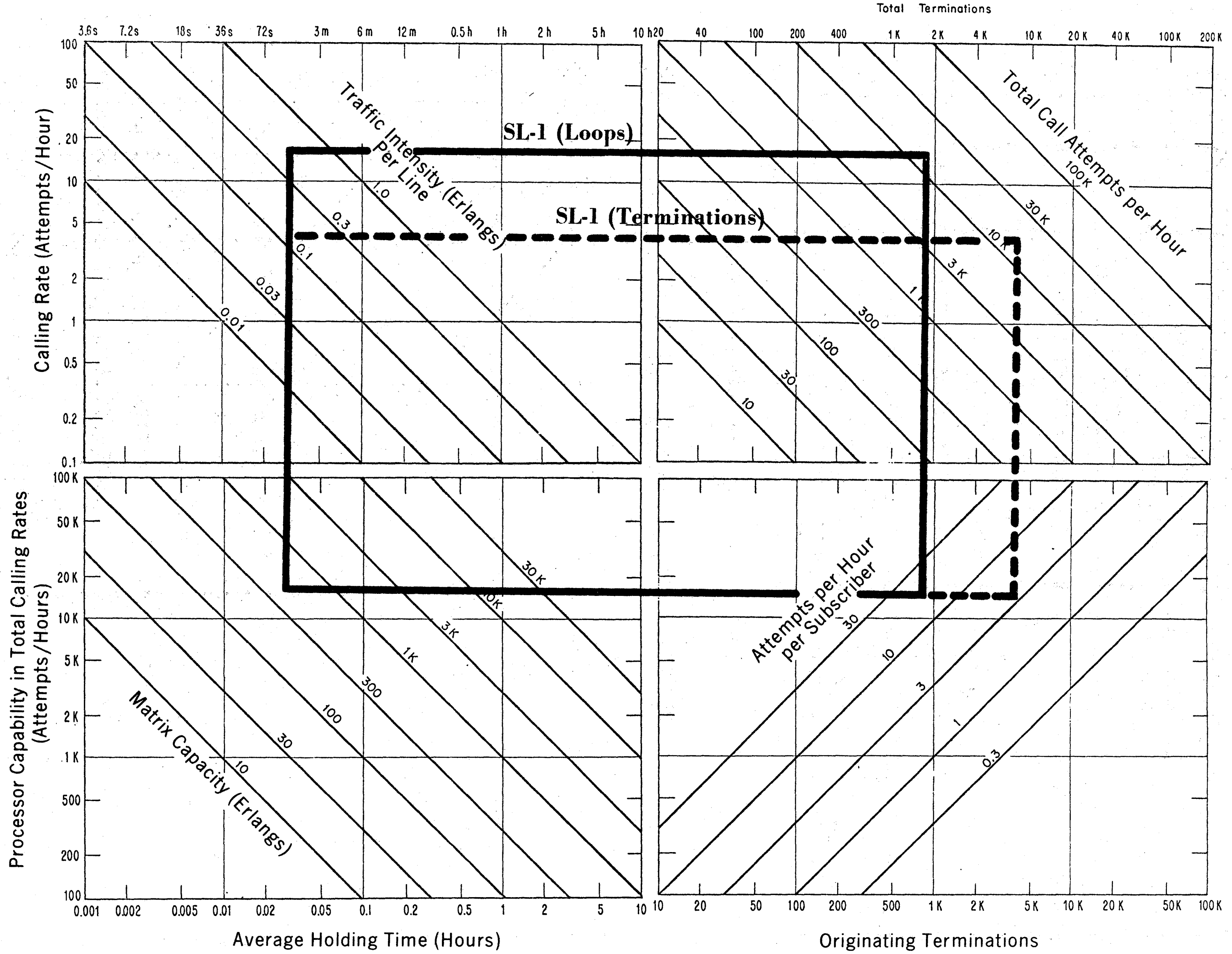


Figure 24. Composite of figures 14, 21, 22 and 23 for switch system capacity evaluation.

Each 30-channel loop is capable of carrying 600 ccs (16.7 Erlangs) of traffic. This is based on a one-way blocking probability of 0.001, assuming blocked calls are cleared, and 0.0023, assuming blocked calls delayed.

The total traffic carried by the switching matrix is therefore

$$\frac{12x600x5}{2} = 18,000 \text{ ccs or } 500 \text{ Erlangs}^{5}.$$

Based on an average holding time of 2 minutes (0.033 hours) per call, the total number of call attempts per hour that must be handled by the processor is given by:

$$\lambda = \frac{A}{\tau} = \frac{500}{.033} = 15,000.$$

The number of originating or receiving attempts per hour per channel is, therefore, 15,000÷900=16.7. The average traffic intensity per individual loop channel is 16.7x0.033=0.55 Erlangs.

The results can be approximated using Figure 24. The solid lines in the figure denote the SL-1(VL) for 5 groups. The dashed lines on Figure 24 are for the same switch, but with a total of 7600 input terminations. The latter concentration ratio is 7600÷1800=4.22. From Figure 24 it can be seen that these 7600 terminations may have 4 call attempts per termination and a traffic intensity of 0.132 Erlangs (4.75 ccs) per line, for an average call of 2 minutes.

The characteristics of the tactical switches given in Tables 5 and 6 provide additional examples of how Figure 24 may be used. The capacity of the TTC-39 switch are summarized as follows:

Call attempts per hour	3,300
Switch matrix load (Erlangs)	180
Terminations	600.

These capacity values are drawn on Figure 25. Completing the square on this figure yields a traffic intensity per termination of 0.6 Erlangs. This corresponds to 11 call attempts per hour per termination with an average holding time per call of about 3.3 minutes. The dashed lines on Figure 25 are for the TTC-38 (300 line version), a tactical switch whose characteristics are given in Table 5. This switch as a space-divided switching matrix and

[5] This is the maximum traffic carried if all loops are full. It may never be achieved due to availability limitations in setting up the required connections through the matrix.

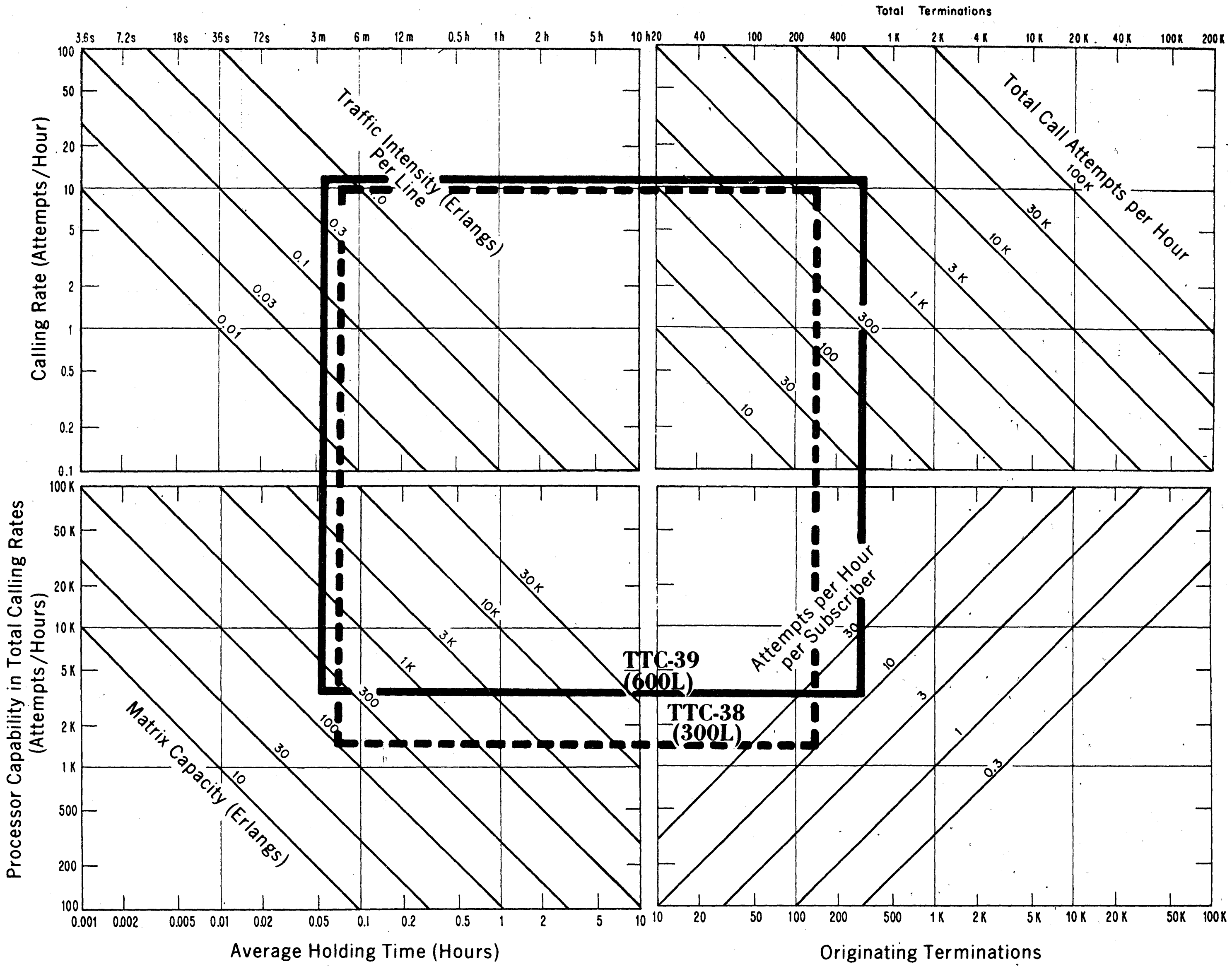


Figure 25. Overlay demonstrating capacity of large toll switch.

uses PNPN diodes for the crosspoints. The basic characteristics are essentially the same as the TTC-39. The TTC-38 processor is rated at 2700 BHCA for a 600 line switch instead of 3,300 BHCA given for the TTC-39. This would result in fewer call attempts per line for a 600 line switch and a somewhat greater holding time. The traffic intensity for both switches is 0.6 E/line. This appears high for an average subscriber's line. However, both switches are designed for use either as an end office switch or a tandem trunk switch. Supposedly they could also be used as a combined end office and tandem switch.

Figure 24 can be further exploited to define and to quantify what one means by such terms as a small, medium, large, and so forth, switch -- in terms of matrix traffic and processor capacity. Several such arbitrarily selected switch size ranges are illustrated graphically in Figure 26 and summarized in Table 8. It is assumed therein that the average holding time per call is 3 minutes.

The general framework for processor capability and matrix capacity limitations, as introduced in Figure 23, may be easily applied to several commercial switches. This is done in Figure 27. The total number of lines and trunks for each switch are also indicated on this figure. The point where the processor capacity and matrix capacity cross is engineered to reflect reasonable balance. That is, the processor capacity need not exceed the switch matrix capabilities by more than a safe margin, and vice versa.

The factors which limit the processor's capacity are discussed in the next section.

4.5. Capacity Limitations of the Control Element

The functions performed by the control processor were listed earlier in Table 2. These functions must be contrasted with the factors that limit the processor's capabilities (see Table 7). In this section these functions and capability limiting factors are discussed in detail.

The processing element of a stored program control must function in real time. The processor's capacity is defined as the highest number of calls processed in a given period of time. This capacity depends on a number of factors, such as the nature of the traffic, the architecture employed, the organization of the stored program, speed of logic, and amount and types of

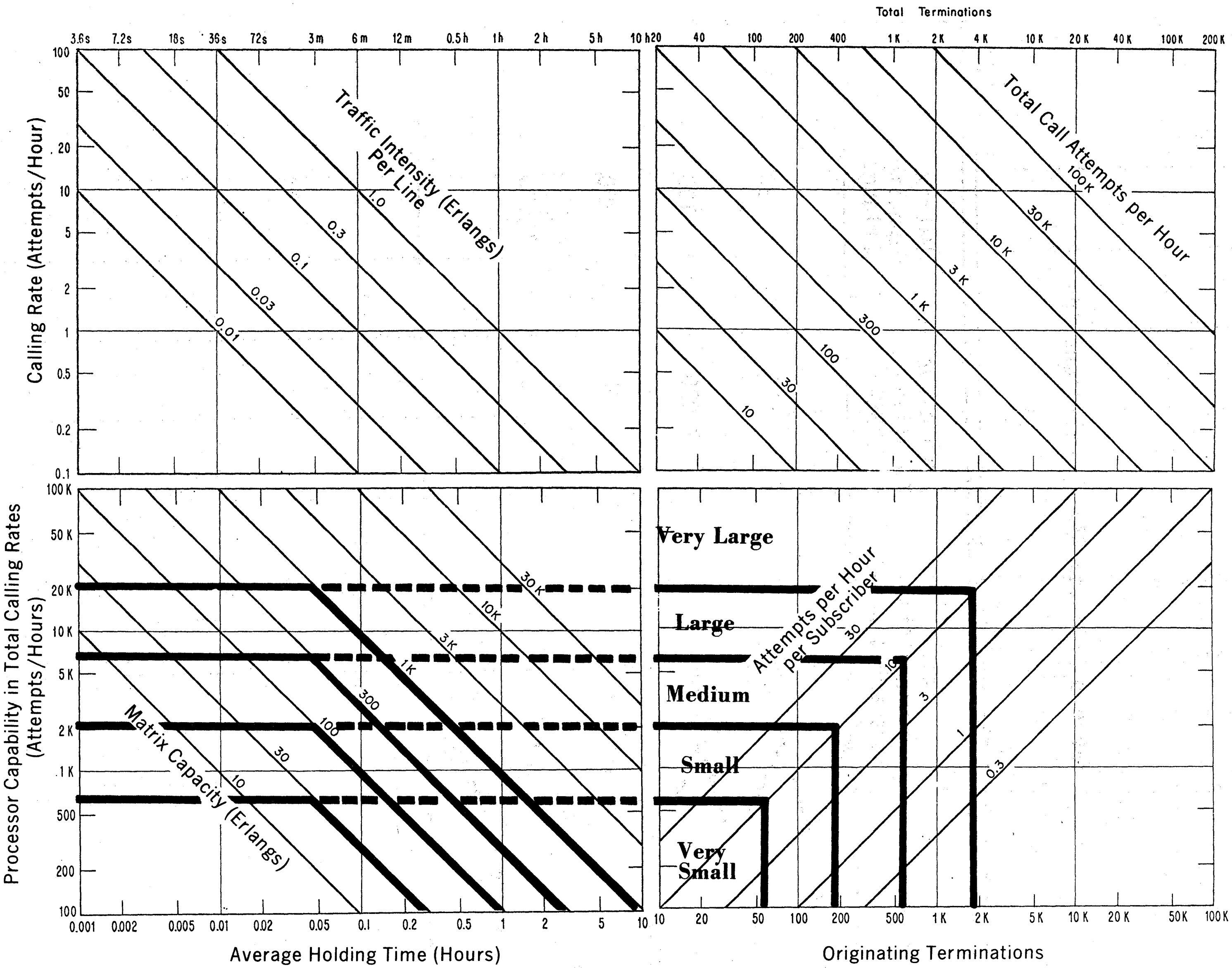


Figure 26. Switch sizing relationships.

Table 8. Switch Sizing Relationships

Size	Total Terminations (lines and trunks)	Processor Capabilities (BHCA)	Matrix Capacity (Erlangs)
Very Small			
	120	600	30
Small			
	400	2,000	100
Medium			
	1,200	6,000	300
Large			
	4,000	20,000	1,000
Very Large			

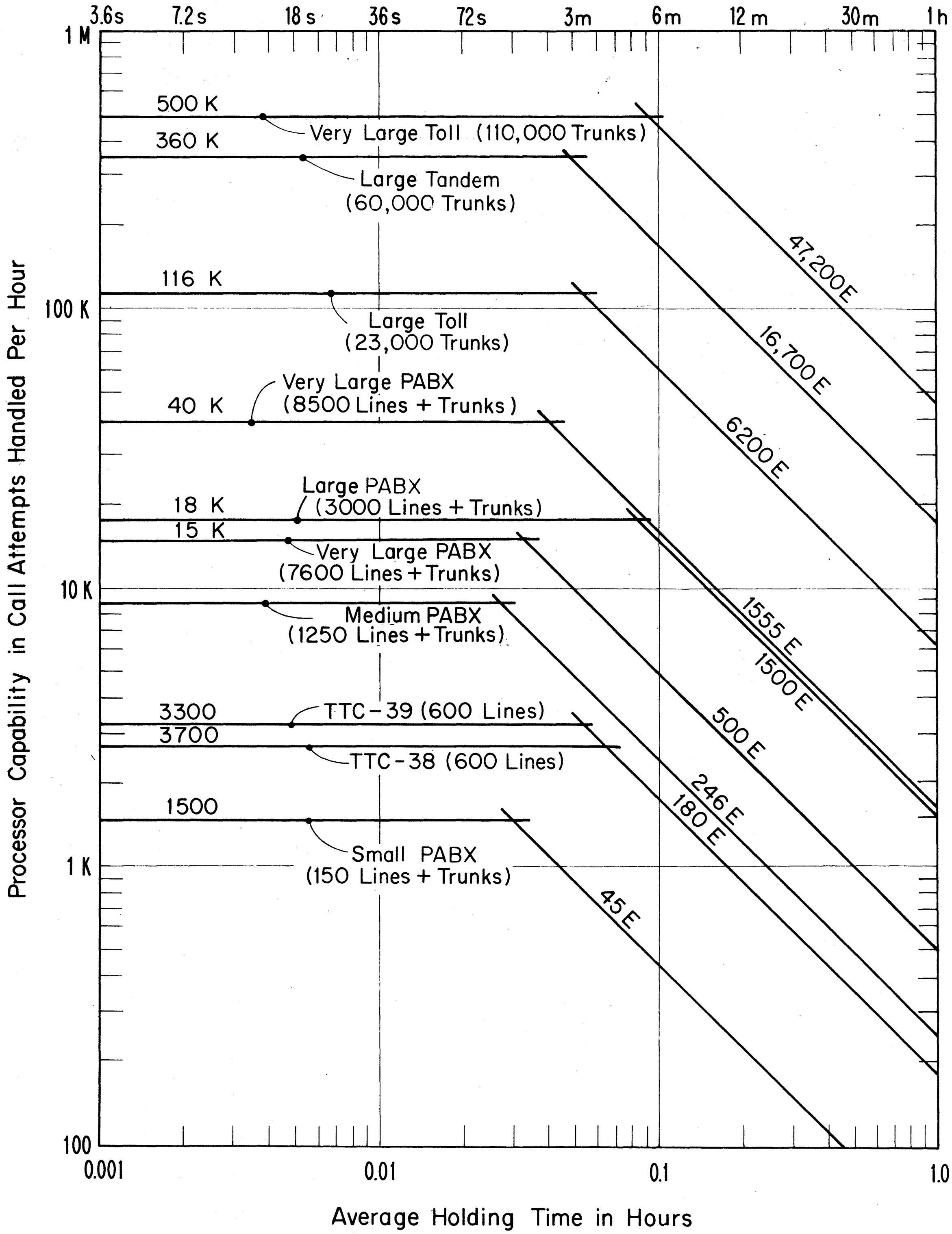


Figure 27. Capacity estimates of some commercial switching systems.

memory. The processor capacity may also be limited by overhead time requirements. An example of overhead is the line scanning function. Scanning must be performed continously. When the number of lines is increased, more scanning overhead is required and less real time remains for call processing. Also, the capacity may be reduced as more service features are offered to the terminals. Thus, abbreviated dialing requires table lookup time. If performed often by the processor, it results in fewer total calls in a given period.

The major factors that limit processor capacity are discussed in the following subsections.

4.5.1. Nature of Traffic

The processor capacity for a given size switch can be estimated using the curves in Figure 24. One convenient method is to determine the number of busy hour call attempts per unit of carried traffic. This, of course, depends on the average holding time per call. Figure 28 is a linear plot of Erlangs per call attempt versus minutes per call. The slope of the line is 0.0167 Erlangs per call attempt per minute. Experience has shown that for each .05 Erlangs of traffic carried by the matrix, there corresponds a processor work load equivalent to one busy hour call attempt. This amounts to an average holding time of 3 minutes (Fig. 28).

Table 9 lists the total number of switch terminations which can be handled by a control processor limited to 10,000 busy hour call attempts. The values indicated can easily be estimated from Figure 24. Table 9 shows, for example, that 10,000 terminations, each carrying 3-minute calls that add up to 0.1 Erlangs, can be handled by the assumed processor. The switch matrix for 10,000 terminations carries 10,000x0.1÷2=500 E (see Sec. 4.3). If the rule of thumb of .05 E per call attempt (see Fig. 28) applies, the processor is required to handle 500/.05=10,000 call attempts, as specified.

Under these conditions, the number of call attempts handled by the processor is comparable to the total number of terminations. Tables 3 and 4 show that this relationship is conservative. This is partly the result of increasing processor capabilities to allow instantaneous peak loads, and partly due to higher loading on trunking facilities.

Some switching systems are engineered with a 30% processor time reserve for peak overload demands (Nowak, 1976). Another 20% may be reserved for

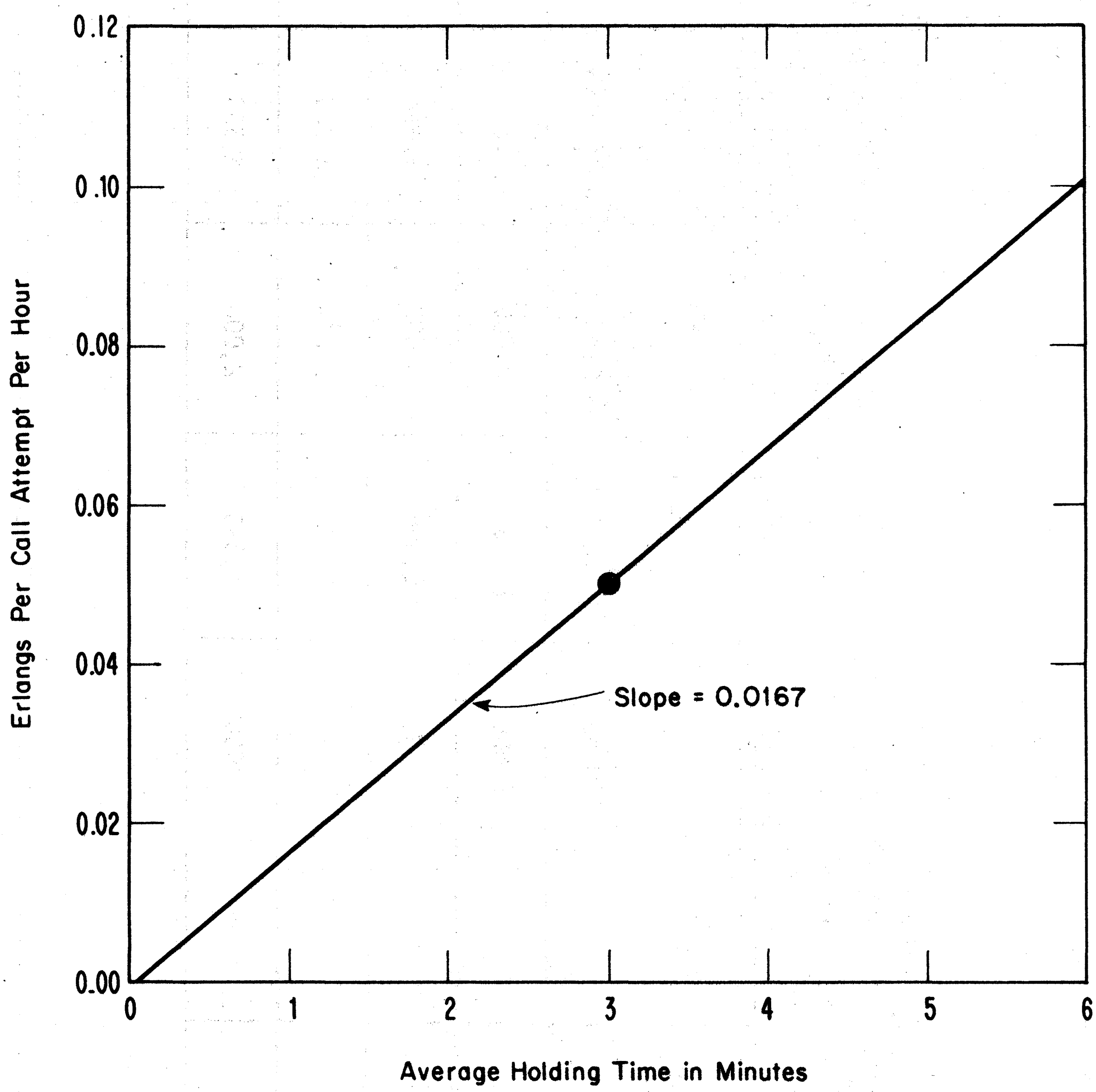


Figure 28. Estimating processor capability requirements.

Table 9. Possible Number of Switch Terminations When Control Processor is Limited to 10,000 Busy Hour Call Attempts

Carried Load per Termination (Erlangs) \ Average Holding Time	1 Minute	2 Minutes	3 Minutes	5 Minutes	10 Minutes
0.05	6,666	13,333	20,000	33,333	66,666
0.10	3,333	6,666	10,000	16,666	33,333
0.20	1,666	3,333	5,000	8,333	16,666
0.50	666	1,333	2,000	3,333	6,666

fixed overhead functions, not related to traffic load. This leaves one-half of the busy hour for call processing. The time available to process each call attempt for a 10,000 termination switch is therefore roughly 180 ms.

A switch of this size may have on the order of 10,000 instructions in the program, of which 60% are devoted to call processing (see Sec. 4.5.3). The average time alloted per instruction is therefore about 30 μs in this exchange.

A large percentage of call processing instructions involve table lookups and updates. A typical table search instruction requires 10 read or write cycles. With the exception of microcycles that consume less than 1 μsec, typical cycles tend to consume 1 to 2 μsec. Thus, the 30 μsec per instruction appears reasonable. Larger programs may be required for special service features. In that case, either the processing speed must be increased or the number of terminations reduced (see Sec. 4.5.5).

4.5.2. Control Architectures

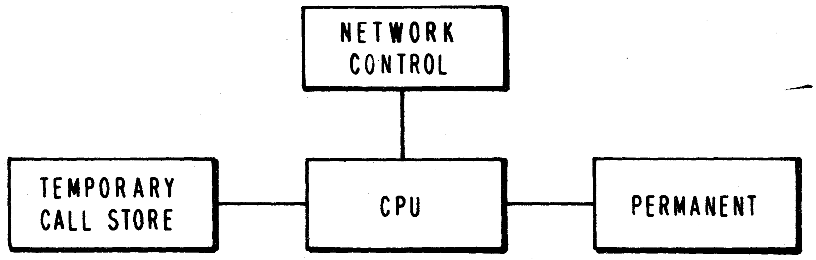
In any circuit switch, the control element performs specific functions based on received signaling information. Call processing functions include recognizing a request for service, registering the directory number, translating these numbers into routing information, and creating a common path between terminals by activating the switching matrix. In addition, the processing element performs maintenance and administrative functions.

In stored program control systems all of these functions are performed by software logic programs and computer processing equipment. These control processors may be implemented using various architectures, as illustrated in Figure 29.

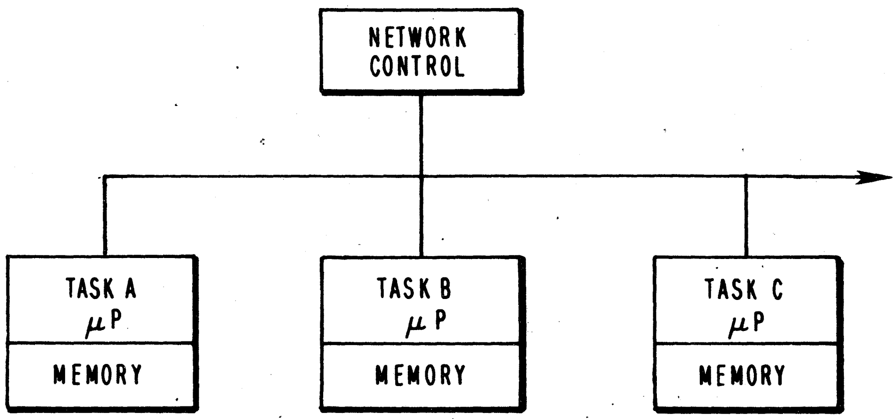
The control processing architectures can be divided into two broad categories, centralized and distributed. In the centralized approach, a single processor performs a multitude of tasks. In the distributed approach, multiple processors perform one or more individually assigned tasks either on a fixed or dynamic allocation basis.

Centralized architectures usually provide a duplicate processor for reliability. This redundant processor may share the work load or be operated on a standby basis.

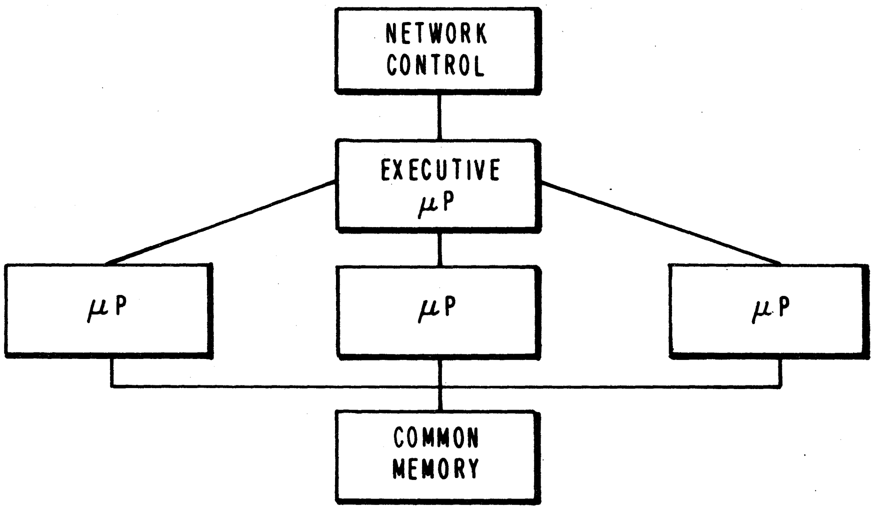
In the distributed architecture, the processing functions are partitioned into tasks that can be handled individually by separate processors. These



a) Centralized Control Element



b) Distributed Control with Fixed Task Allocation



c) Distributed Control with Dynamic Task Allocation

Figure 29. Possible control element architectures for stored program controlled switches.

processors may share a common memory or have their own associated memories, such as in the multicomputer approach. Multicomputer systems often have one microprocessor which assigns tasks to the others.

Distributed systems, that operate with a fixed allocation of tasks, often incorporate separate processors for reliability: either one for one, or one for many. Dynamic assignment architectures require less redundancy to achieve the reliability.

The centralized architecture is commonly used in both older and larger switching systems. As the size is reduced, the larger more centralized controls tend to increase the cost per line. Distributed architectures may then be more cost effective. Just where the break point occurs is not clear. Distributed architectures have been used in systems with a few thousand terminations. Other benefits attributed to distributed architectures include increased call handling capability and a reduction in the amount of information transfer between switch elements. A possible disadvantage is the typical need for more software development.

4.5.3. Software Organization

In an SPC switch, the software programs can be divided into four basic categories, as shown in Table 10. The number of instructions in each category depends on the type of switch and the features and functions offered. Almost half of the total program may be used for diagnosing faults in peripheral equipment, trunk or service circuits, the processor itself, and for miscellaneous test purposes. Some of these diagnostic programs may be stored in an off-line device, such as magnetic tape, and loaded when needed.

The operating system programs provide task scheduling and control of the input/output interfaces. The application programs provide call handling, while the administrative programs provide system support in terms of program modifications, data updates and accounting functions.

The total number of instructions/program for some commercial switches are plotted in Figure 30. The data points represent different kinds of switches with different word sizes. In general, the number of instructions/program increases with increasing number of terminations. This is expected for the diagnostic portion. The application program size varies with the number of features and functions offered. Usually, larger systems offer more service features and administrative functions.

Table 10. Software Categories

Operating System Program

- o Task scheduling
- o Peripheral control
- o Work coordination

Application Program

- o Call processing
- o Traffic flow control

Administrative Program

- o Software modifications
- o Data base updating
- o Accounting
- o System monitoring

Diagnostic Programs

- o Fault detection
- o Maintenance
- o Miscellaneous tests

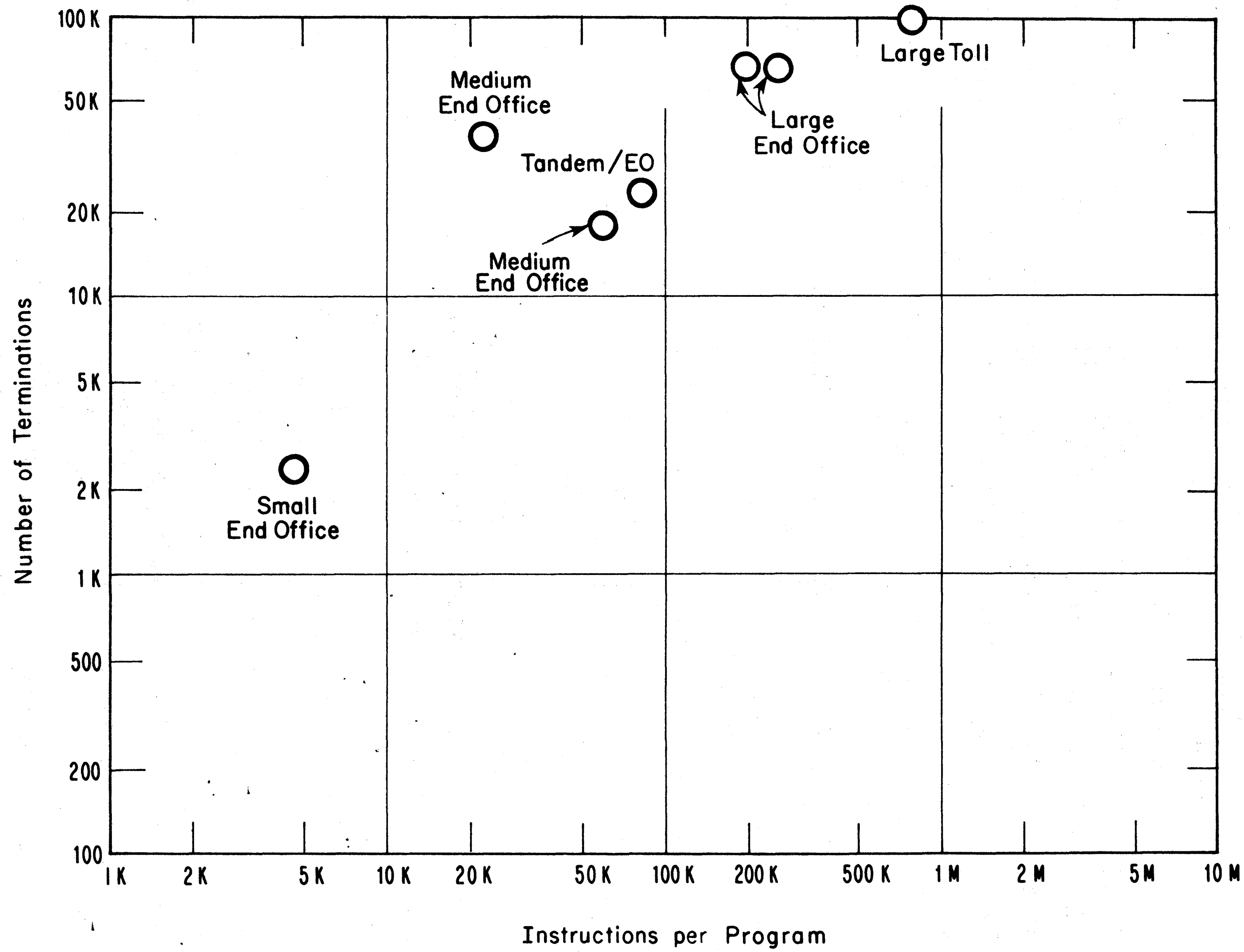


Figure 30. Program size for some commercial switches.

In the past, the programs for most switching systems have been written in assembly language. The software development effort for a typical switch family may be broken down roughly as follows:

Analysis and Design	30%
Coding	25%
Debugging	45%

The program development for one SPC system (No. 1 ESS) is summarized in Table 11 (Harr, 1968). The first two lines of this table are on-line programs prepared in assembly language. The number of debugged words per man-year is probably somewhat low. A more typical value is on the order of 1000 words or less per man-year.

Table 11. Software Development for No. 1 ESS

Software	Programmers	Program Words	Man-years	Words per Man-year
Operation	83	52K	101	515
Maintenance	60	51K	81	630
Compiler	9	38K	17	2230
Translation	13	25K	11	2270

The compiler and translation programs in Table 11 are off-line programs and may have been partially written with a higher level language. The use of such compilers results in increased productivity per programmer (perhaps as high as four times).

The use of higher level languages for on-line programs appears to be evolving slowly. The number of debugged words per man-year for high-level languages is probably about the same as for assembly language. However, the program productivity increases since fewer coded instructions are needed per program function. When the high level code is compiled to assembly language, additional overhead occurs. This transition tends to increase the operating time as much as 30%. It is the main drawback for using higher level languages.

Many recent basic concepts and techniques, that have developed for programming computer-controlled switching systems, are described by Hill and Kano (1976).

4.5.4. Memory Size

The processor memory size tends to increase as their capability increases. This is indicated in Figure 31. The memory requirements in words, versus number of call attempts per hour, are plotted for a selected number of commercial switches. The points must be properly interpreted, since the word size varies between certain switches. In general, the word storage tends to be ten times the processor capacity (in call attempts/hour).

4.5.5. Service Features

In a stored program controlled switching system, the processor's activities can be divided into three task areas:

1. Continuous, constant rate overhead tasks which are independent of traffic; e.g., line and trunk scanning.
2. Intermittent call processing tasks which occur randomly, but increase with traffic; e.g., address translation and path selection.
3. Deferrable tasks, which can be accomplished during off-peak traffic periods; e.g., routine maintenance and administrative tasks.

The manner in which these tasks are scheduled impacts the demands on processor capacity. The number of call processing tasks and the time required to perform each task ultimately sets a limit on the processor's capability. This limit may be defined as the maximum number of calls which can be processed in some fixed interval of time while meeting all the performance objectives (Brand and Warner, 1977). A major consumer of processor capacity is the service feature set offered to the customer. The number and kinds of service features available is one of the advantages of SPC switching. Features can easily be added or changed by program modifications via attendant consoles. Special station apparatus may be necessary to take full advantage of the more advanced features. Table 12 lists some of the basic and enhanced services listed in a typical manufacturer's sales brochure.

In the military switching environment, additional software packages could be added to provide new services, such as precedence, preemption, and restricted access to assigned groups of users. Security and privacy may also be required for specific calls.

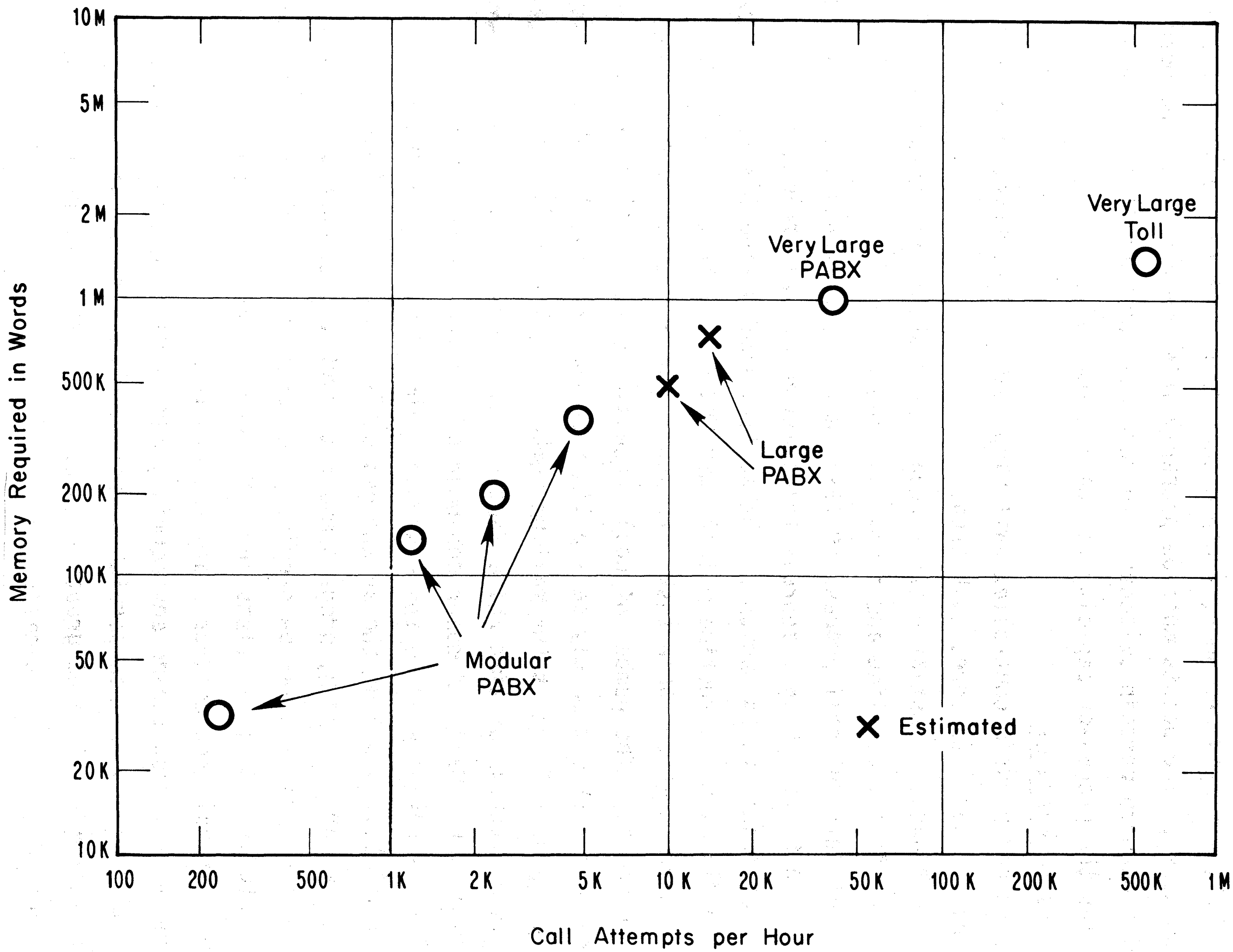


Figure 31. Memory requirements versus processor capability.

Table 12. Selected PABX Service Features

Basic Features (Standard Apparatus)	Additional Features (Special Apparatus)
o Ring again	o Call waiting
o Call transfer	o Call forward
o Conference	o Manual intercom
o Call waiting	o Abbreviated dial
o Call pick-up	o Automatic dial
	o Extension override

Special feature requirements can affect the processor's capability. The call processing time increases as features are added, thereby reducing the number of call attempts handled in a given period. The capacity limit of the switch depends on the mix of station apparatus and special features employed. This is illustrated by Figure 32 where the total processor activity time is divided into overhead time, which is independent of the traffic load, call processing time (which depends on traffic load), and maintenance time. The fixed or recurrent overhead time depends on the type of switch. The overhead may vary depending on the particular application of the switch. In a given application, the processor must continuously perform certain functions such as scanning the lines and trunks, and monitoring status regardless of the traffic load. As the traffic load increases, call processing activity time exceeds the basic overhead time. At lower traffic, deferable tasks can be performed to fill the available time. Ultimately, as more and more calls arrive, the processing time increases to a point where handling of additional calls becomes impossible. Usually, such switch congestion violates some service criteria (e.g., dial tone delay). The traffic then has exceeded the capacity limit. This limit can vary by as much as a factor of two or more, as the mix between standard and special features changes. Addition of special features to a given switching system reduces the number of terminations the switch can handle.

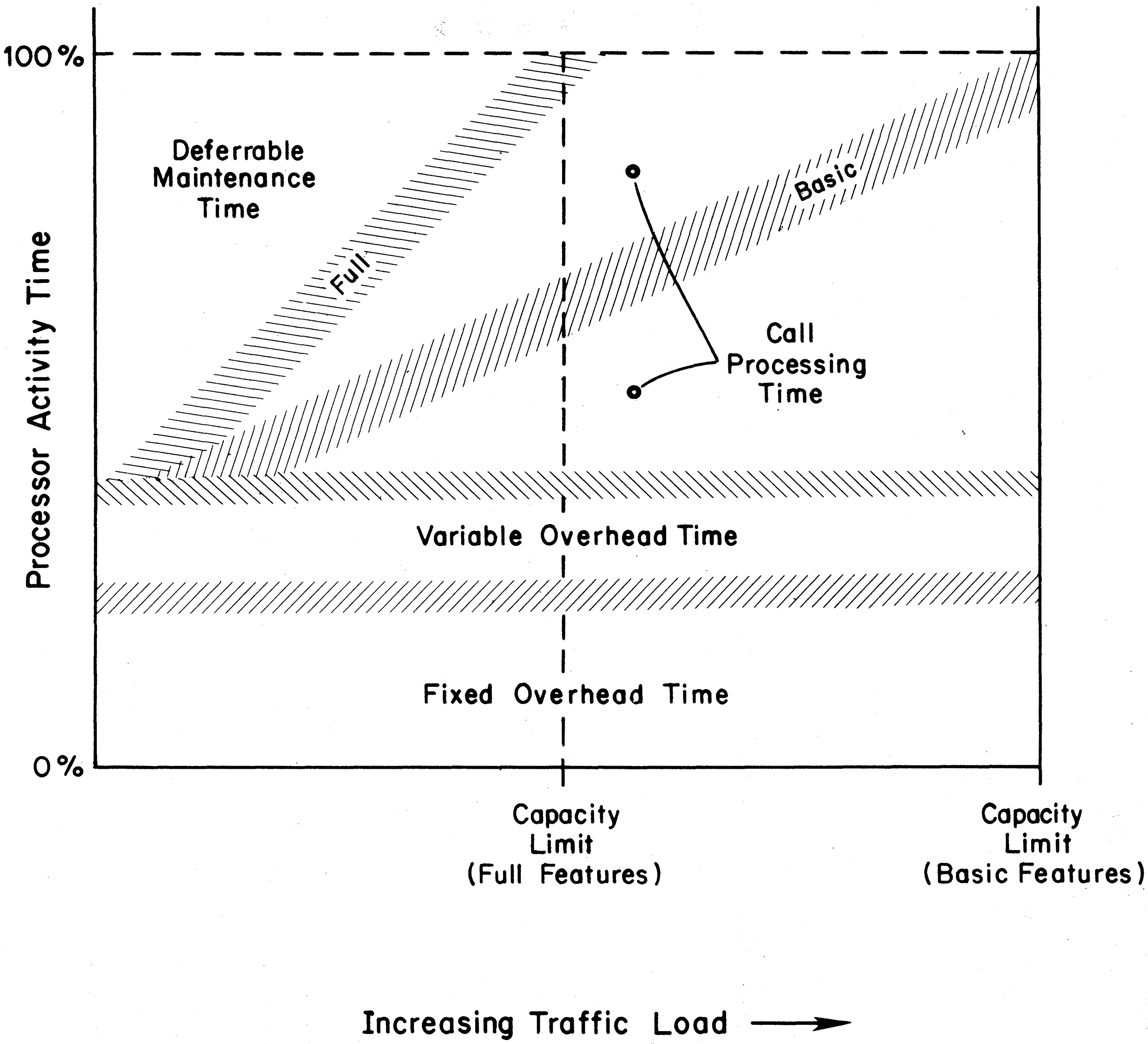


Figure 32. Impact of service features on processor capacity.

4.6. Control Signaling Systems' Impact on Capacity

Signaling systems provide the means for remotely controling the switching system. In this section, one is concerned with signaling techniques used on both the line and the trunk sides of the switch. Station signaling on the line side and interswitch signaling on the trunk side may take various forms. There may be in-band or out-of-band signaling on analog circuits. These may be in-slot or out-of-slot signaling on digital circuits. A review of various control systems which exist today or are contemplated in the near future is given by Linfield (1979). In this section, the impact of some commonly used signaling systems on the capacity of a SPC switching system is discussed.

4.6.1. Station Signaling

The call handling capacity of an SPC switch varies with the types of station apparatus, the call mix and, as noted previously, the service features offered.

The station apparatus types include rotary dialing or dual tone multiple frequency (DTMF) dialing.

The DTMF dialer uses 10 out of 12 standard push buttons to send combinations of address digits, 0 through 9, using a code consisting of two in-band frequencies emitted simultaneously. The additional buttons may be used for special services, such as priority calling. The time for the dialing operation increases by about a factor of two over DTMF when a rotary dial telephone is used. The rotary dialer emits a controlled sequence of dc pulses on the subscriber's line. The number of pulses between breaks corresponds to the decimal digit, except that 10 pulses are used for the 0 digit.

Table 13 summarizes the time factors involved in accessing the telephone network. The connect time increases for interswitch calls, because the processing times for each switch in tandem are additive.

A number of tone generators, registers and senders from a common pool are usually shared at the switch. On the average, that decreases the total switching time. Incoming calls are held in memory by the processor until the appropriate common equipment becomes available. This queueing for service equipment provides some smoothing of the offered load bursts. It permits the processor to handle simultaneous calls in sequence, and increases the effective processor capacity.

Table 13. Representative Values of Contributors to Telephone Network Access Time

<u>Dial tone delay</u>	
Light traffic	0.1 s to 0.5 s
Heavy traffic	0.1 s to 100 s
<u>Dialing time</u>	
Rotary	10 s manual
DTMF	4 s manual, 0.7 s automatic
<u>Connect time</u>	
Local calls	1 s to 5 s
Trunk calls	10 s to 15 s

The average amount of time per call required by the processor also depends on the call mix offered by the stations. Call mix varies as a function of local subscriber population. Nominal call percentages at a commercial switch are as follows:

Completed calls	60%
Busy and no answer	20%
Improper dialing	15%
Other (includes blocking)	5%.

The number and types of service features requested also impact the time per call, as noted in Section 4.5.5.

The station type, call mix and features offered all jointly determine the number of instruction cycles that must be executed to process a call. This was shown earlier in Figure 32. The average time per call increases as features are added, decreases as DTMF telephones are added.

4.6.2. Interswitch Signaling

Interswitch signaling provides communications between processors in SPC switches. In the past, this has commonly been accomplished on a per-trunk basis. More recently, common-channel signaling systems of various forms have been introduced (e.g., CCIS, CCITT No. 6, CCITT No. 7) for use on public switched telephone networks. Common-channel signaling techniques combine the signal requirements for several channels onto a single separated signaling channel. Common-channel signaling can also be used between remote concentration points and the switch and on multiplexed buses within the switch itself.

Since common-channel interswitch signaling (CCIS) has several advantages over per-trunk or per-channel signaling, this technique will be discussed next. Emphasis here will be on its impact on switch system capacity.

A block diagram of a common-channel interswitch signaling system is shown in Figure 33. The control information to be transferred between processors is usually delivered "in parallel" to the signaling terminal. The information is formatted into blocks called signaling units (SU's), with different SU's coded for different types of information. Check bits are added for error control. Figure 34 illustrates several SU's used by the CCIS system of North America. Each SU contains 20 information bits and 8 check bits. The CCIS steps involved in a successful completion of a 10-digit call are shown in Figure 35.

The traffic intensity of the common signaling channel, A_s, depends on the number of trunks being serviced, the average trunk intensity, A, the average signaling message length, ℓ (in bits), and the signaling rate, c, in bits/sec.

The signal channel intensity can be expressed as

$$A_s = \lambda_s \, \ell/c,$$

where ℓ/c is the average holding time per signaling message and λ_s is the arrival rate of the message. If each SU is 28 bits in length (see Fig. 34) and an initial call set up requires 6 SU's, then the average holding time per call on a 2400 b/s signaling channel is

$$\frac{\ell}{c} = \frac{6x28}{2400} = 0.07 \text{ s}.$$

Assume that the largest trunk group contains 10 channels and each channel carries 0.7 Erlangs of traffic (0.35 in each direction) with an average holding time of 3 minutes. Then the total number of call attempts to be handled by the signaling system is given by

$$\lambda_s = 10 \times \frac{0.7}{180} = 0.0388 \text{ attempts/s}.$$

The signal channel intensity for 10 trunks is therefore

$$A_s = \lambda_s x \ell/c = .0388x.07 = .00272 \text{ Erlangs}.$$

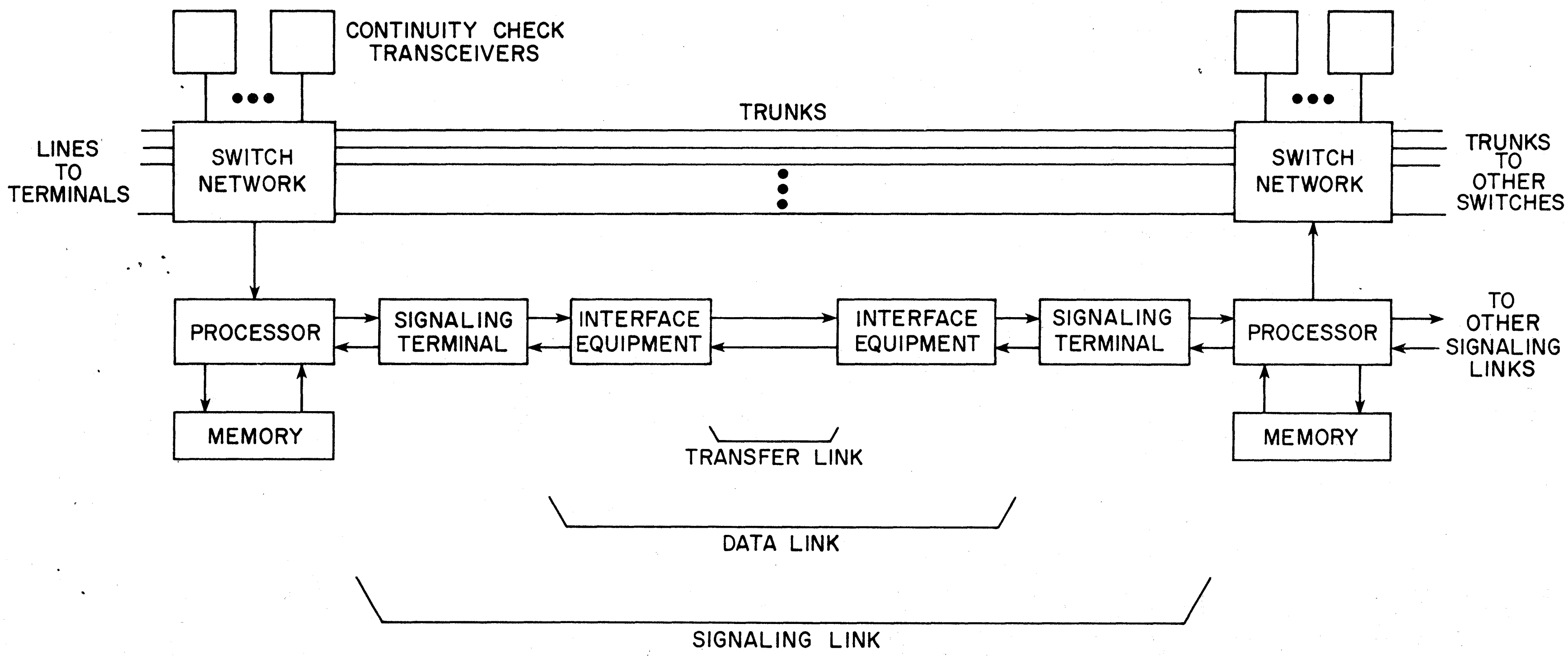


Figure 33. Control signaling via common channel signaling link.

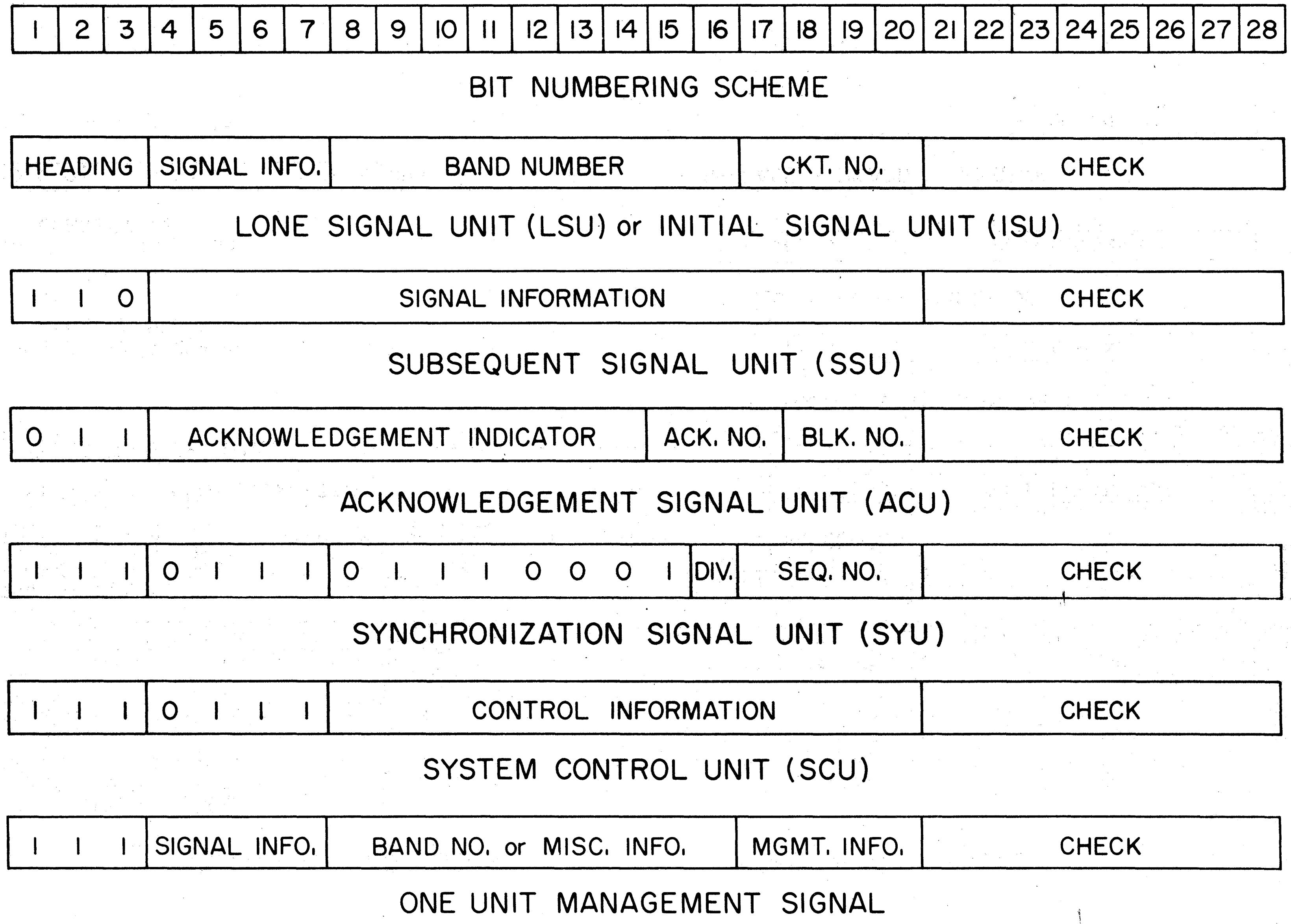


Figure 34. Basic format of signal units used with common channel interoffice signaling system.

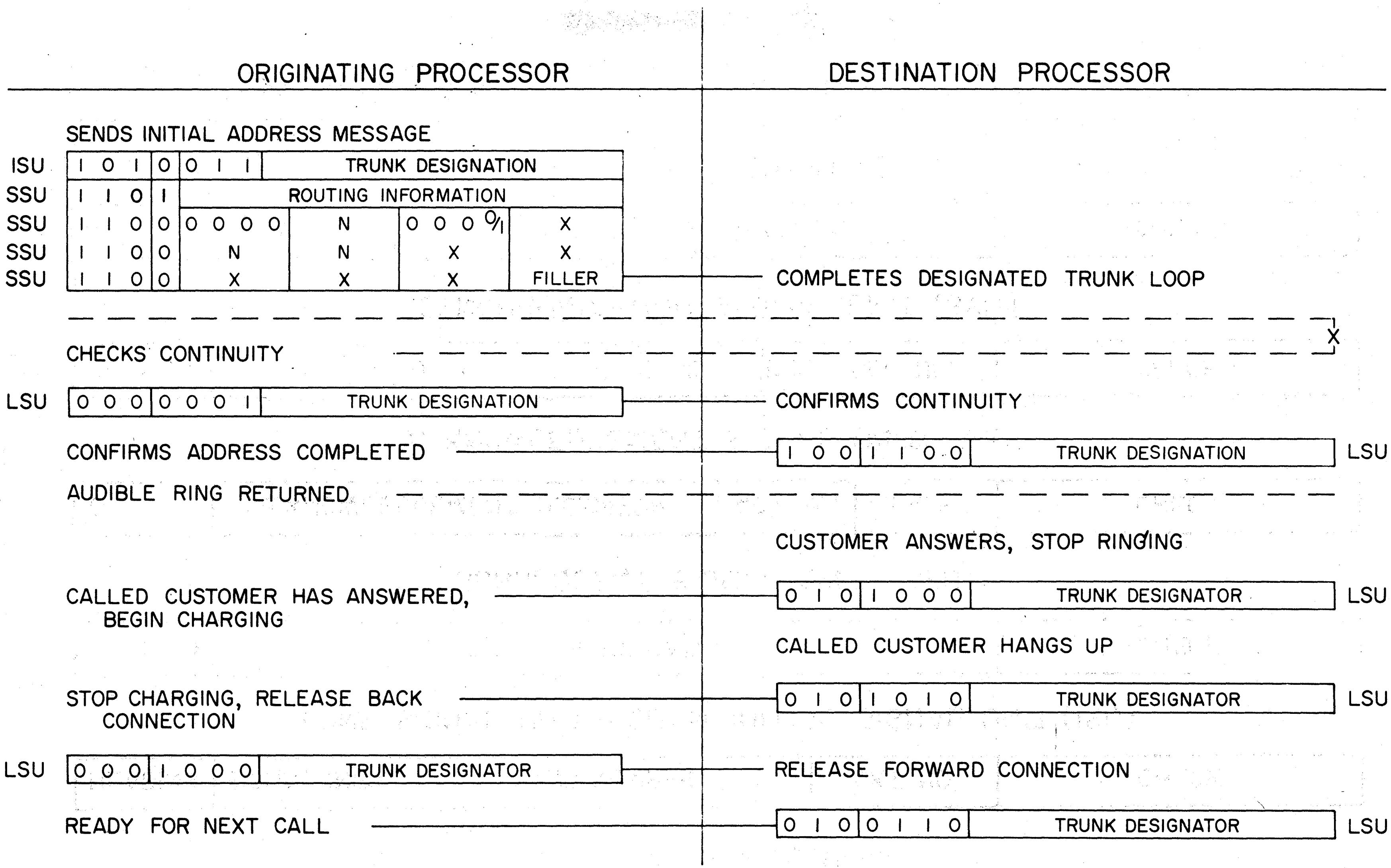


Figure 35. Steps required for 10-digit call using CCIS.

A single signaling channel may carry 0.01 Erlangs with a blocking probability of 0.01, according to Erlang B blocked calls cleared grade of service tables. The processor, moreover, can hold arriving calls in memory until a signaling channel becomes available. Under these favorable conditions, each signaling channel can be packed to carry up to 1 Erlang. The 2400 b/s CCIS channel used in North America is designed to handle 1500 trunks. Under the conditions stated above, the 2400 b/s CCIS channel carries approximately 0.4 Erlangs, if there are 6 SU's per message, and 0.8 Erlangs if there are 12 SU's per message.

The traffic intensity on the signaling channel for 10, 100 and 1000 trunks is plotted in Figure 36 for 6 and 12 SU's.

Signaling system capacity does not limit the switch system capacity. This is shown in Figure 37, where the number of call attempts offered to the processor, via the common signaling channel, are plotted versus the channel signaling rate and parametric in carried load. Even when the signaling channel is limited to 0.1 Erlangs, the number of call attempts is over 80,000 for a 64 kb/s signaling rate. This is the standard rate available in one digital PCM voice channel.

5. TRAFFIC ENGINEERING

5.1. Needs and Background

The objective of this section is to review traffic efficient design and use of access area networks. The section deals with common configurations of user lines, trunks, concentrators, switches, and other service subsystems that perform the communication and switching tasks required.

Emphasis is placed on the networks' abilities to carry projected telecommunication traffic volumes. Clearly, it is cost prohibitive to engineer systems meant to handle infinite amounts of traffic. Systems are always costwise and physically limited, and so is their ability to handle extreme surges of traffic. Network performance under specified traffic scenarios is nominally described in terms of "grade of service" (AT&T, 1961; Siemens, 1970; Linfield and Nesenbergs, 1978). Unless otherwise specified, grade of

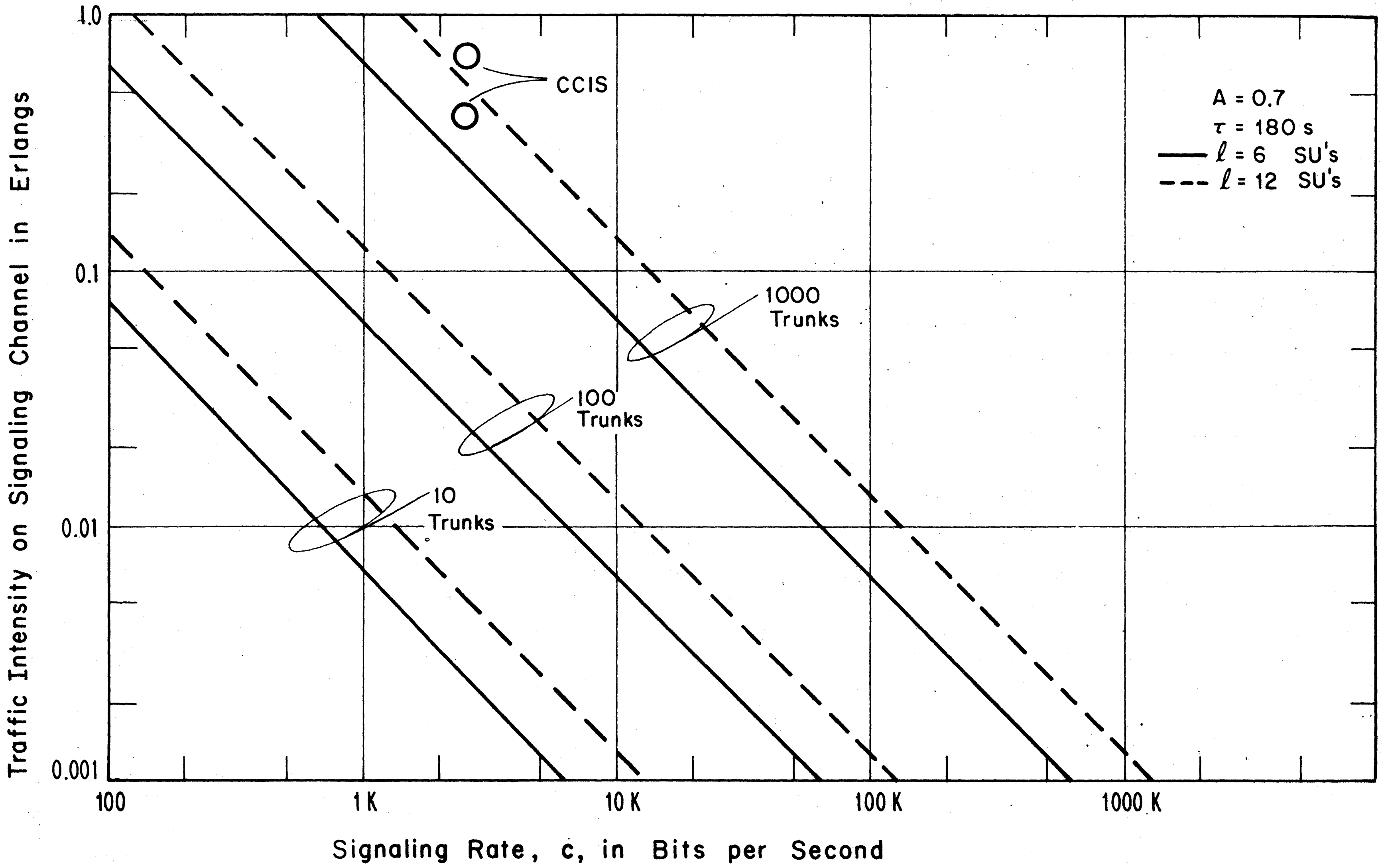


Figure 36. Common channel signaling traffic intensity versus signaling rate.

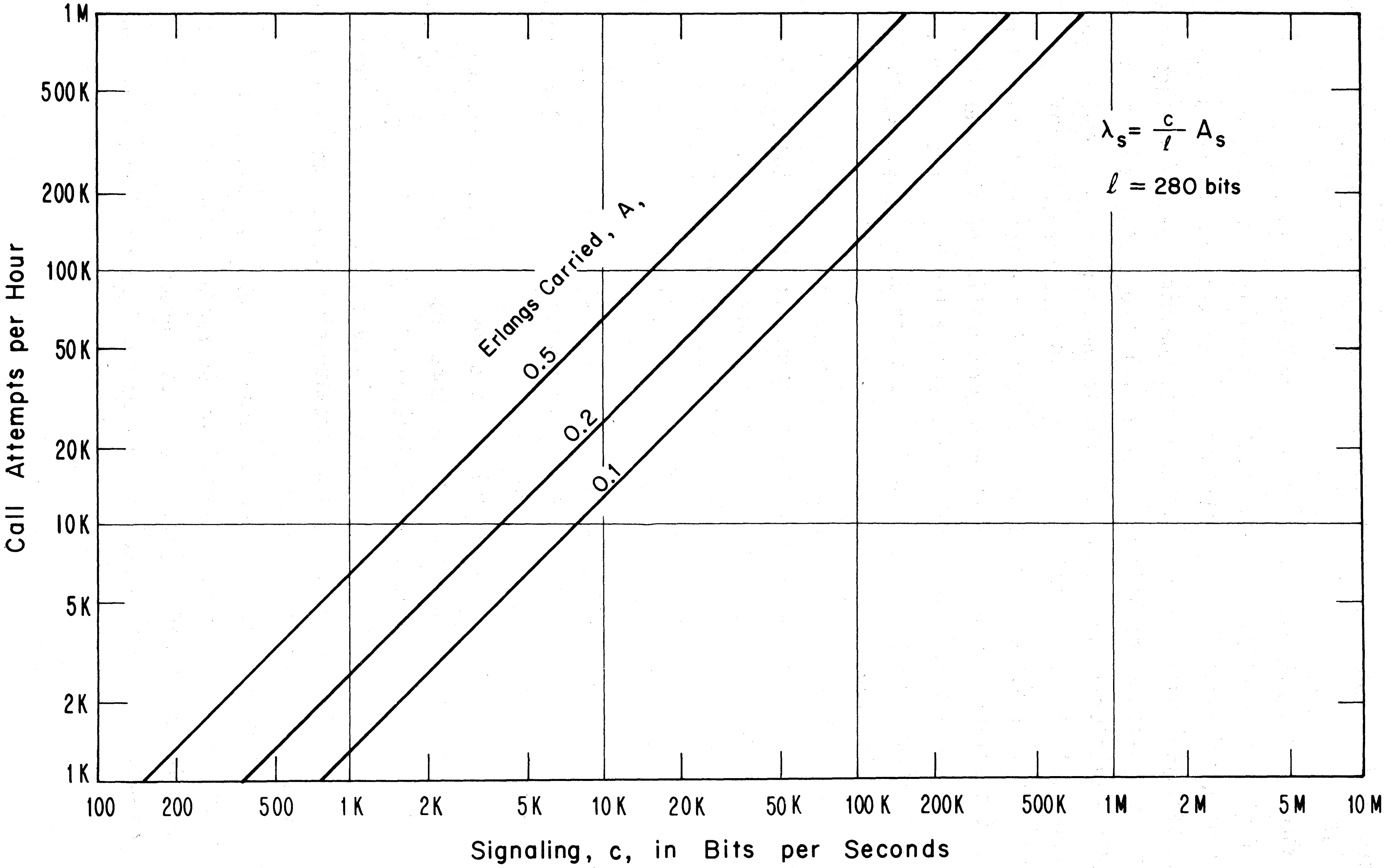


Figure 37. Common channel call attempt capacity.

service refers here to probability of blocking. A typical objective for commercial and military grade of service may be a blocking probability on the order of 1%.

When service requests (i.e., calls) encounter blocking, several things may happen. The blocked calls may be dropped or rerouted (cleared), and thus lost from the service facility in question. Or the blocked calls may be queued (delayed) until served. In general, queueing systems include random delays, possibilities of buffer overflow, and a host of tradeoffs that far exceed the simpler grade of service concept and scope (Fry, 1928; Morse, 1958; Saaty, 1961; Kleinrock, 1975; Bear, 1976; and Schwartz, 1977).

Because of space and time limitations, this section will touch delay systems but briefly. Emphasis will be on probability of blocking effects for postulated access area subsystems. A summary of the more relevant communications and switching aspects for both the loss and delay system types is presented in Table 14.

Let us explain the terms used in the table. A service system or facility may be first characterized by its general type. It may be called loss, delay, or some loss-delay hybrid, depending on what happens to typical blocked service requests. The distribution of service request initiation times or arrival times plays a major part in determining whether a service process is at all analytically tractable or not. The Poisson distribution of arrivals per unit time is equivalent to exponential, memoryless, or Markov (M) interarrival distribution. It is an ideal model that is extremely helpful in any analytical thrust at traffic engineering.

The distribution of service or holding times may also affect the system performance. Availability refers to the number (or fraction) of serving trunks which can be accessed by an individual offering trunk (Fry, 1928; AT&T, 1961; Siemens, 1970). It is simpler to deal with switches, PABX's, data buses, distribution modules, and concentration modules that are arranged to provide full availability.

When faced with delay systems, the order of service, such as first-in-first-out (or FIFO, or strict queueing) versus random order, affects the delay statistics of specific service attempts. Preemptive (i.e., interruptive) or non-preemptive priority structures may be required by access area users.

Table 14. Methods of System Characterization

System Descriptors	Typical or Tractable Models
General Type	Loss Delay Loss-Delay Hybrid
Arrival Process	Poisson (Exponential) Overflow (Peaked)
Service (Holding) Time	Exponential Deterministic (Constant)
Availability	Full Limited
Order of Service	Order of Arrival (FIFO) Random Specified Priority
Random Delays	Probability Distribution Mean, Variance Limited (Bounded)
Buffer Storage	Infinite Finite
Reroute Provisions	Fixed (Including None) Adaptive
Users or Sources	Infinite Population Finite Population
Servers per Facility	Single Multiple

If delayed, the duration of delay is typically a random variable whose distribution may be hard to ascertain. The mean or the variance may be more tractable, however.

Near the conclusion of Table 14 one finds three numerical descriptors. The first, the amount of buffer storage, is usually weighed against transmission and switching capabilities. Joint minimization of costs and service delays is the issue here. The second number mentioned is the user or source population size. If it is infinite, one may be justified to use either the Erlang B or the C formula, but only if the system fits the rest of the Erlang model (Fry, 1928; Morse, 1958; AT&T, 1961). If the number of sources is finite, grade of service may be found from the Engset results (Riordan, 1962), that assume blocked calls cleared and finite number of sources as part of the system model. Tables and charts are available for both Erlang B and C, as well as for the Engset formula (Siemens, 1970; CCITT, 1977b). The tabulations are classified in accordance with the number of servers, such as the number of access ports, serving trunks, or the equivalent TDM time slots.

A broad look at the "now" status of telecommunication traffic engineering reveals the following. First, the bulk of present R&D activity in this country and abroad concerns the delay, waiting time, and other queueing aspects of data oriented networks (Kleinrock, 1976b; Schwartz, 1977). This is indeed a fast developing field, spurred by packet switching, and rich in problems and solutions. Second, the older interests in grade of service (or probability of blocking) seem to have settled down on a more stable and comfortably accepted level. This, of course, does not mean that all grade of service questions are either trivial or readily answered.

Just as in the twenties, when T.C. Fry postulated his eleven General Assumptions (Fry, 1928), classical blocking concepts are still valid for all circuit switched, voice and data, systems. Several new and useful tools have appeared during the last quarter century. The present situation faced by grade of service estimators is outlined in Figure 38.

The tree chart of Figure 38 shows that exponential arrival and service (holding) time distributions are by far the most popular and manageable premises for traffic engineers. Likewise, infinite sources, full network availability, and systems of the plain loss type (i.e., with blocked calls

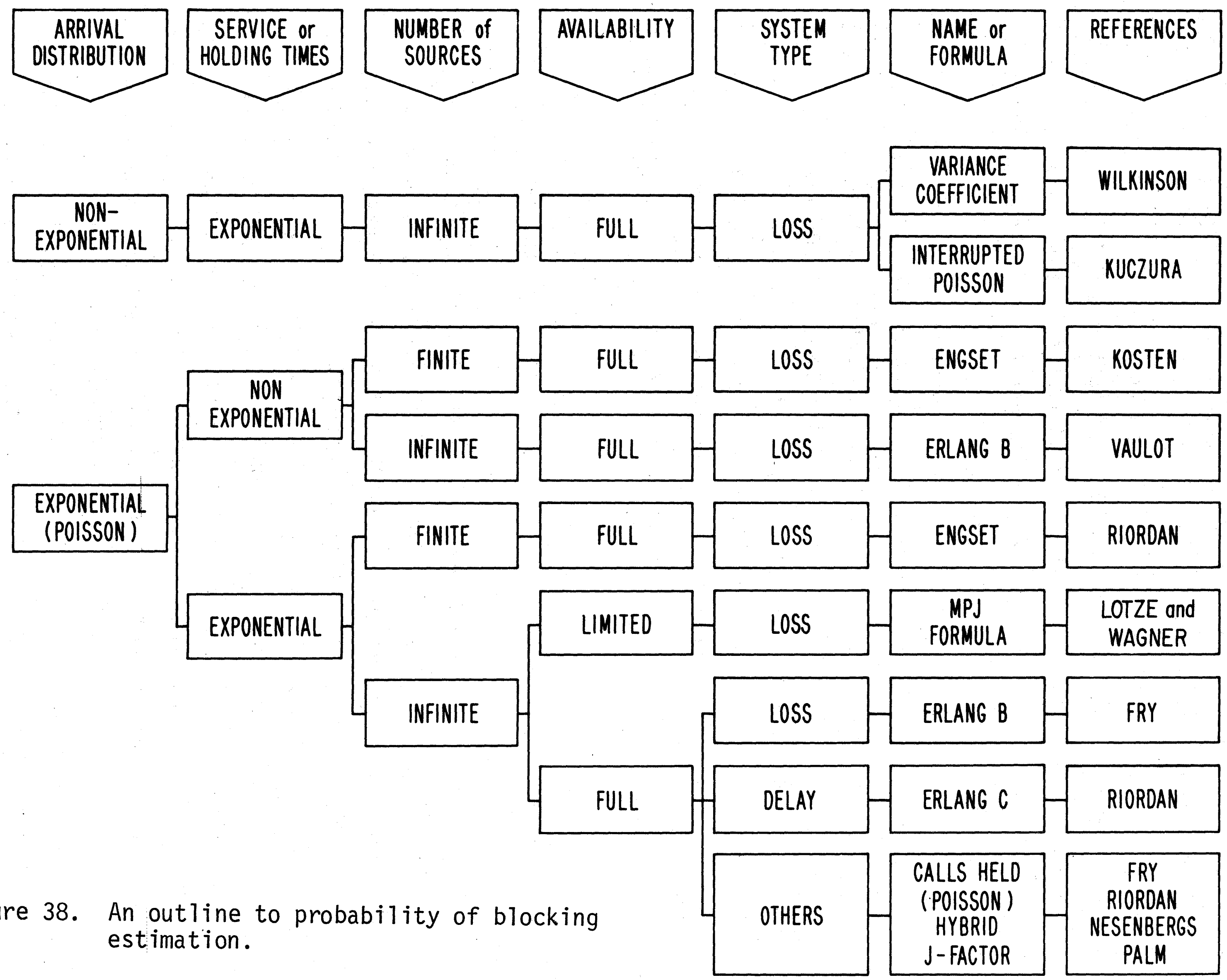


Figure 38. An outline to probability of blocking estimation.

lost or cleared) are the easiest to handle. Erlang B, C, Engset, and even Poisson blocking probability formulas are still quite useful. New variants, such as the Modified-Palm-Jacobaeus (MPJ) formulas for limited availability, have been tabulated (Siemens, 1970).

The basic or source references are listed on the right side of Figure 38. Several of them, such as Kosten (1948-1949), Vaulot (1927), Lotze and Wagner (1963) or Palm (1937), may not be readily accessible. The work of Kosten, Vaulot and Palm are summarized in Riordan (1962). The, so called, MPJ method is described in Siemens (1970).

Yet, with all these analytical tools, not everything needed is readily available to both military and commercial telecommunications traffic engineers. For instance, there appears to be considerable uncertainty about the efficient planning of server trunks or local tie-lines between a concentrator (or multiplexer) and its nearest switching node. It is also difficult to depict adequately the effects of two widely differing user sets (e.g., lines and trunks) on a common m-server facility. These two topics will be discussed in the following two sections. After that, several other areas of unresolved traffic issues will be indicated.

5.2. The Server Trunks Between a Concentrator and Its Switch

The efficient use of tie-lines (or server trunks) between remote or colocated concentrators or multiplexers (MUX) and their nearest switches, such as end offices or PABX's, is important to military access area design because of the number of lines and tie-lines involved. The nature of the problem and the notation is introduced in Figure 39.

Let M users or sources have access to the high-side of a concentrator (i.e., to the side with the most ports). On its low-side, let the concentrator have m server trunks. Typically $M>m$. The m trunks tie the concentrator to the switch as shown. By assumption, let there be no switching at the concentrator. That means that a local, turnaround, call cannot be cross-patched at the concentrator, but must travel to the switch and back.

The facility shown in Figure 39 is seen to have two types of service requests. The calls that are destined to a distant switch require a single server trunk. The local calls, on the other hand, require two server trunks.

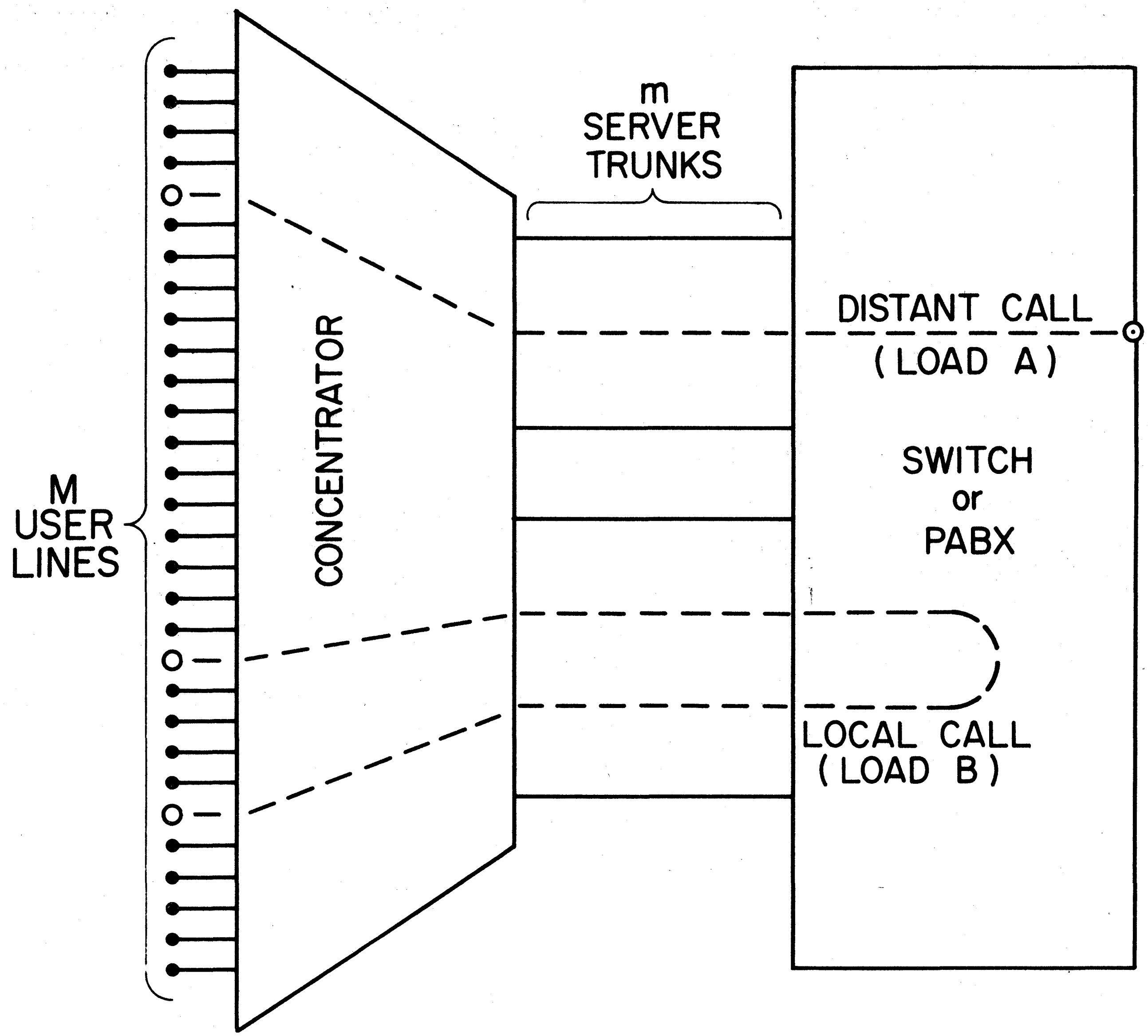


Figure 39. The server trunks or tie-lines between a concentrator and a switch carrying two types of calls: distant and local.

The grade of service, which appears to be different for the distant and the local calls, depends on the relative frequencies or mixes of the two types of calls.

To proceed, define the following quantities. Let:

a = Distant call load (or intensity) per line

b = Local call load per line

A = Effective total distant load

B = Effective total local load.

Then, as shown in Figure 39, loads a and A are associated with distant calls and require one of the m servers. Loads b and B require a pair of servers.

The key question here is to relate the probabilities of blocking (let P_D stand for the blocking of distant calls and P_L for local calls) to the above loads and to the numbers M and m of users and trunks, respectively.

Without reroute options, the network in Figure 39 is a loss system. If M is finite, one may consider using Engset tables (Siemens, 1970), but there is a question of what constitutes the effective load, and what should be the number of users and servers. The interested reader is directed to the Appendix A where this problem is treated in detail. The dilemma is outlined in Figure 40.

Figure 40 shows the total effective load, A+B, versus m, the number of server trunks, for given M users and at the projected probability of distant call blocking of $P_D=10^{-2}$. The problem is the following. There appears to exist various blocking possibilities, depending on the relative values of A and B. Least blocking occurs when all calls are distant (B=0), most blocking when no calls are distant (A=0). And between these two bounds one has a large region of uncertainty. Without specifying the A/B ratios, a total load of A+B=5 Erlangs may require from as few as m=10, to as many as m=33, server trunks.

Appendix A attacks this problem in considerable detail. While an exact solution is shown to be difficult, approximations are possible at a cost of a reasonable computational effort. Typical results are illustrated in Figure 41. This figure assumes that the distant call load is one-third of the local calls or one-fourth of the total. It also supposes that the probability of blocking based on traffic observations for distant calls is $P_D=.01$. Note:

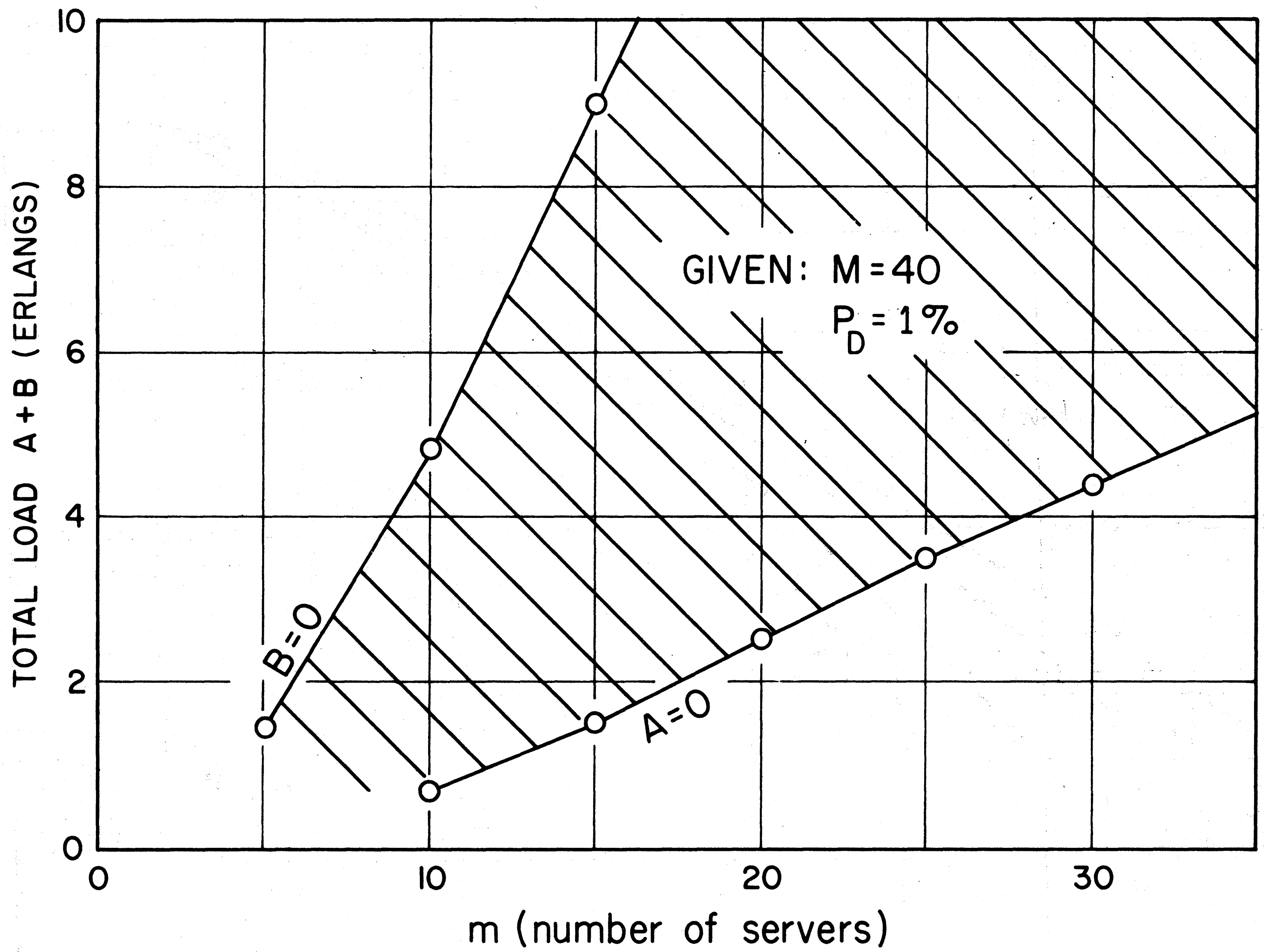


Figure 40. Region of uncertainty created by unknown mix of distant (A) and local (B) call loads.

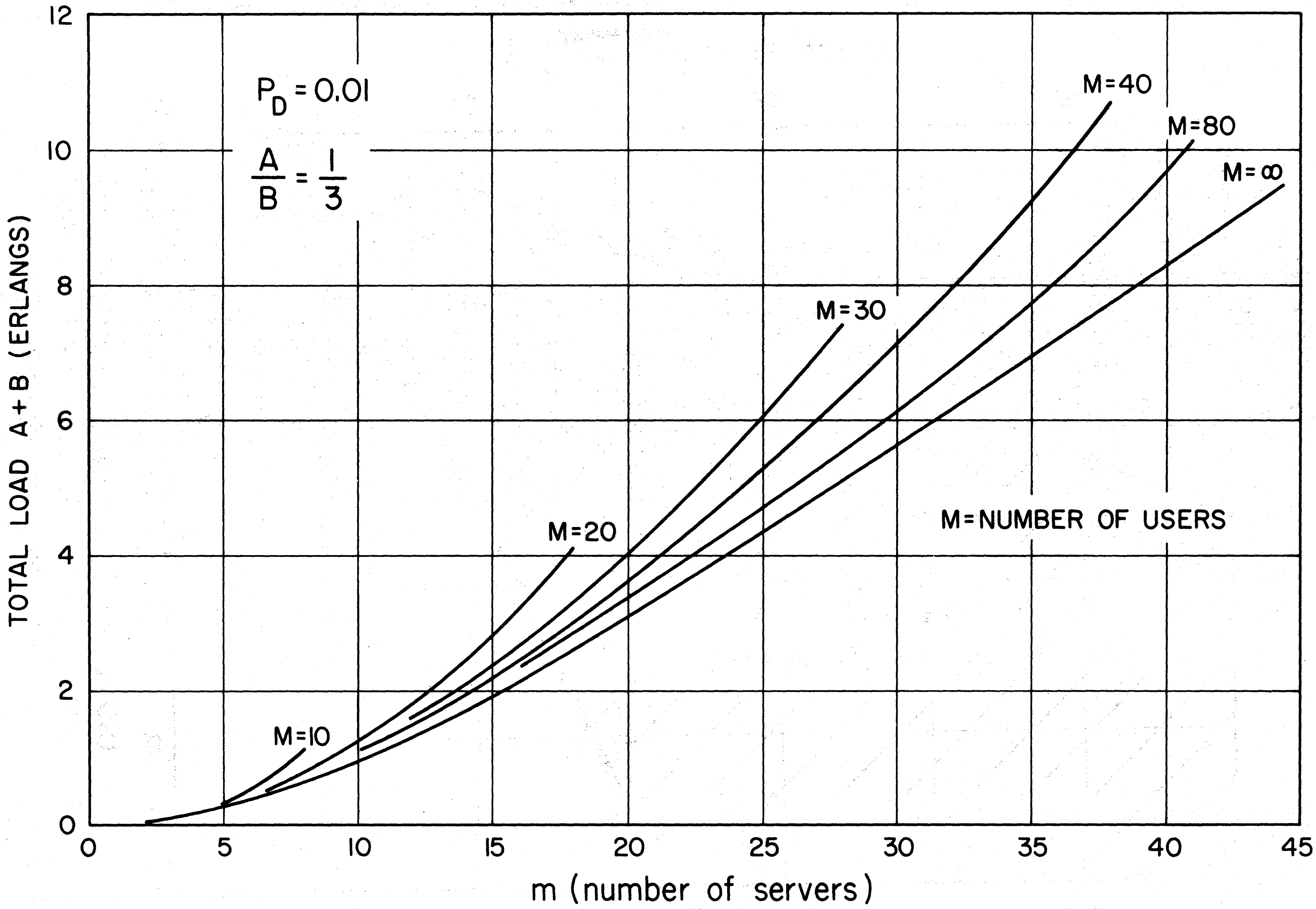


Figure 41. Estimated total load vs. number of servers for the assumed practical P_D and A/B values.

The blocking probability, P_L, for local calls is more than .02 for these assumptions, as described later in Appendix A.

Figure 41 plots individual curves for the number of users, M=10, 20, 30, 40, 80, and ∞. Each curve shows the load A+B achievable with the number of servers, m, given in the abscissa. For example, m=24 server trunks can carry up to A+B=5 Erlangs, total load, if M=40, P_D=.01, and A/B=1/3. Comparison of Figures 40 and 41 reveals that the M=40 curve falls slightly below the middle of the region of uncertainty.

5.3. Common Facility for Two User Types

Flexibility and compatibility objectives of access area subsystems may call for more than one set of users (sources) to interact with a given server arrangement. For instance, several CPU access ports may serve the entirety of programmed PABX actions. There may be more than one memory unit with parallel provisions for read/write by different user groupings. Both hardware and software may often be viewed as an m-server facility common to diverse users.

Figure 42 illustrates three instances of a common m-server facility serving two apparently dissimilar user sets. In part (A), a blocking switch or its main concentration module is fed by L "lines" that come directly from user terminals and T "trunks" from a remote line concentration or multiplexing (MUX) unit. In applications, this wording could be readily reversed. Thus, the L's could be trunk and line ports and the T's could be MUX loops. The main point is that these two user classes request service from a common bank of m ports, perhaps outgoing toll trunks. Typically, there may be more direct station lines (i.e., L>T), but their relative loads are apt to be lighter. Part (B) of Figure 42 shows a common pool of m signaling units, or any peripheral equipment, that serves the control needs of L lines on the local side and T trunks on the toll side. Part (C) of Figure 42 depicts the recently proposed SENET system model (Vena and Coviello, 1975; Fischer and Harris, 1976). This model proposes to integrate circuit switching and packet switching at a common multiplexing or concentration facility that is intelligent enough to assign m channels to said traffic.

One should observe that the three cases of Figure 42 may involve quite different service and traffic statistics. For instance, in (A) the service

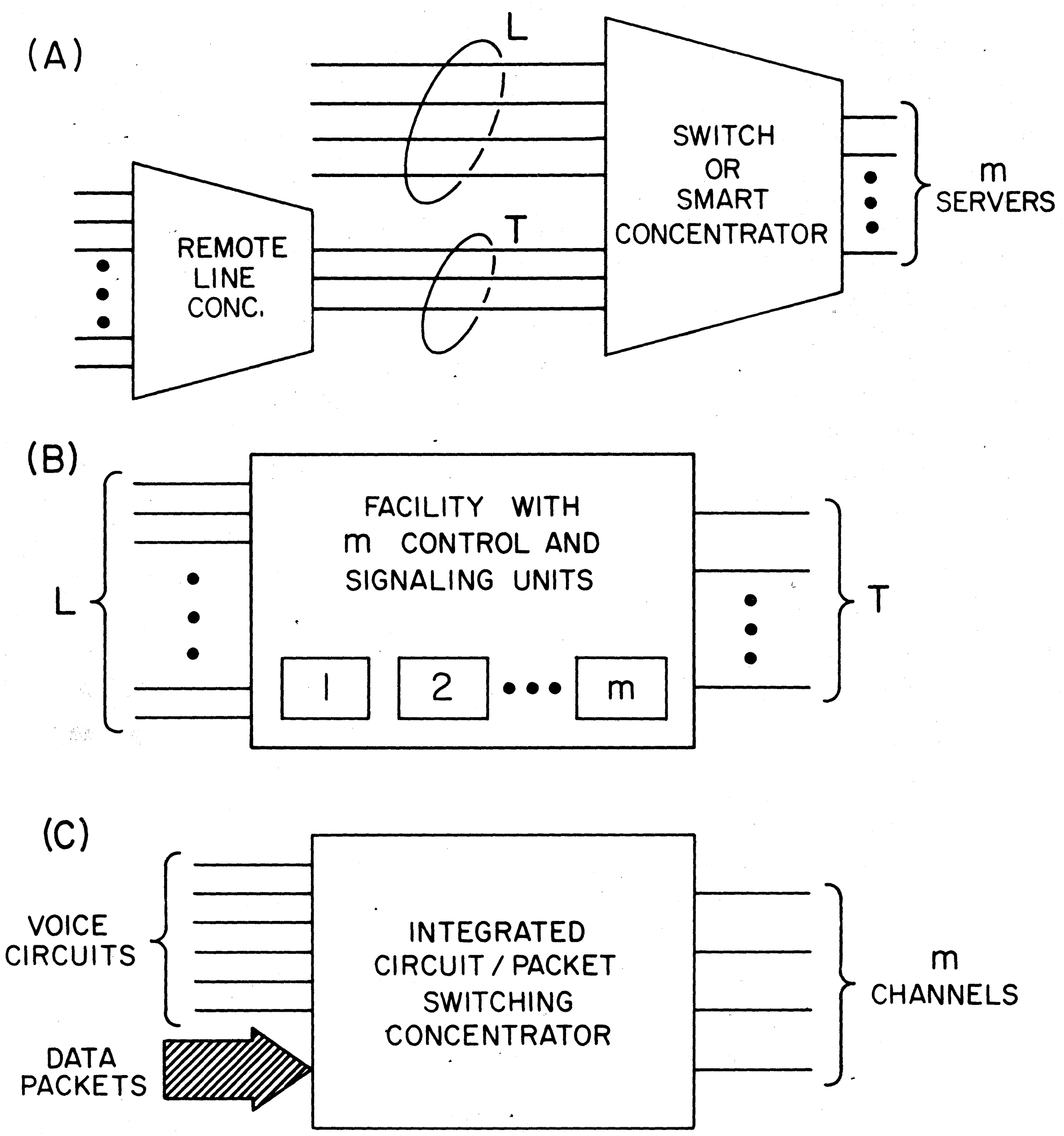


Figure 42. Examples of m-server facilities for two user types: (a) concentration of lines and trunks; (b) sharing of signaling units; and (c) circuit and packet switching integration.

(holding) times may be the same for both classes. In part (C), however, the duration of data packets may be an infinitesimal fraction of a typical phone call holding time. In part (B) the service times may or may not be the same for the lines (L) and trunks (T).

Clearly, calls originate at different rates over station lines and interoffice trunks. Data packets, as in part (C), have their own traffic generation statistics (Kleinrock, 1976b). There are also issues concerning blocked service requests. Certain categories of calls can be justifiably dropped, cleared, rerouted, or plainly lost. Others may be queued, such as perhaps in part (B). And yet others, such as voice lines in (C), may have preemption priority over data. The general topic extends beyond the scope of this discussion.

One must stress nevertheless that this, to our knowledge largely unexplored, queueing model could be broadly interpreted and variously applied to AADSS. The m servers could be tone or ring generators, a bay of operator consoles, billing and accounting machines, route selectors, not to mention old-fashioned markers or futuristic secure voice encryptors/decryptors, secure data units, stored subroutine lookup and action implementation elements [such as always needed for various call forwarding, transfers, abbreviated dialing, conferencing, priority preempting, and other generally or selectively required AADSS features and functions (Linfield and Nesenbergs, 1978)].

This section comments briefly on the grade of service behavior of m servers that carry the traffic of two discernible sets of users. The model incorporates a realistic loss-delay hybrid feature that enables individual service requests to be either cleared or queued when all m servers happen to be busy. The general nature of the model is seen from Figure 43. As shown, the load per line is a. The load per trunk is b. In typical applications one envisions $a<b$ and $L>T>m$, but this need not be a strict universal rule. For simplicity, assume full availability to all m servers from all L lines and from all T trunks.

The definitions of the loss-delay hybrid has been given elsewhere (Nesenbergs, 1979). Whenever all m servers are busy - and immediate service is blocked - the contending user has a choice. He can either join a queue or he can drop out. To make things simple, both types of users, namely both

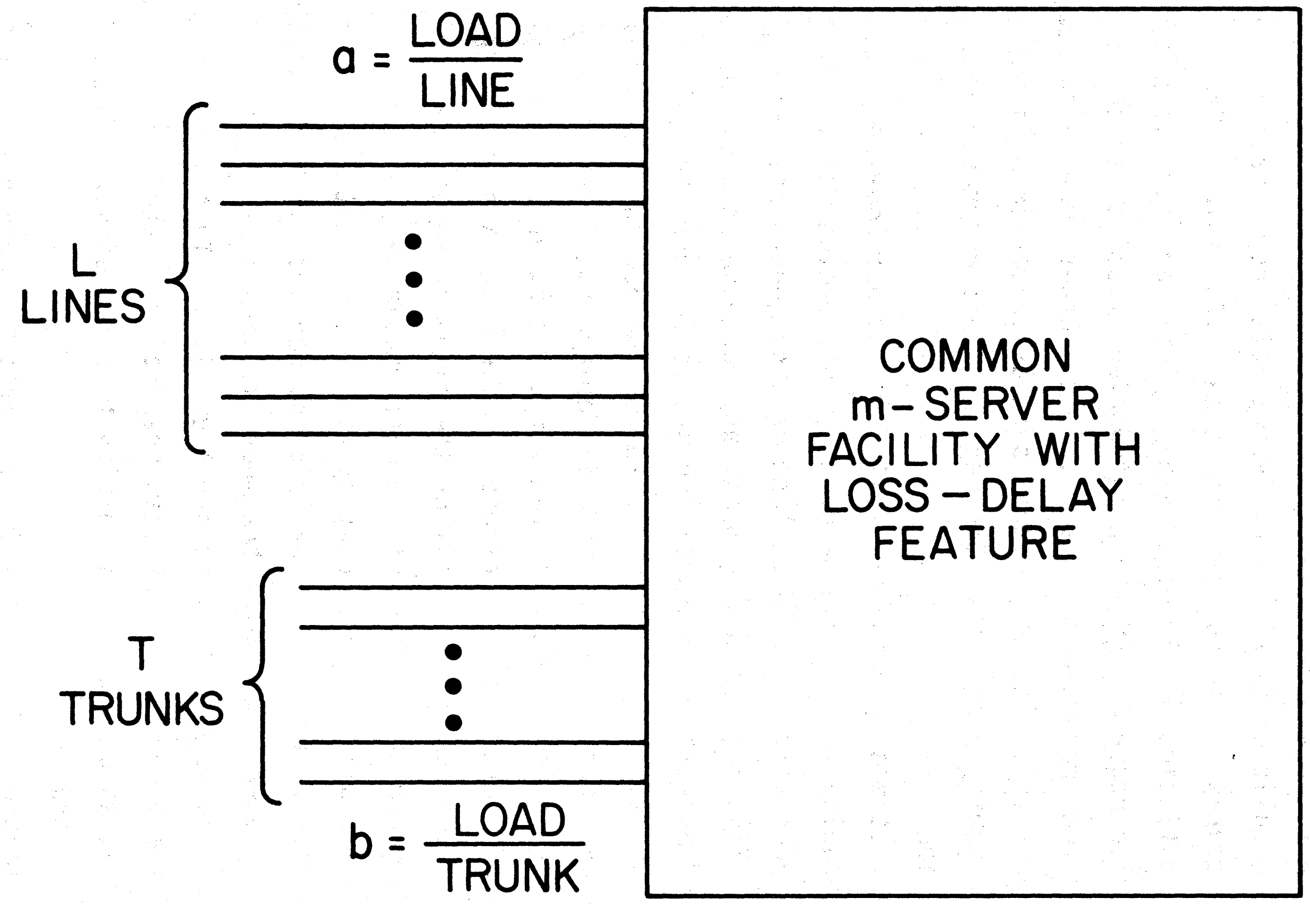


Figure 43. Two user types, called lines and trunks, share a common m-server facility with a loss-delay feature.

lines and trunks, will be assumed to behave in about the same statistical fashion. That is, given blocking, all requests join a queue (and wait until served) with probability θ. They drop out with probability $1-\theta$ and are lost (cleared) as far as the common facility is concerned. The random traffic flow model of the loss-delay hybrid is summarized in Figure 44. In the very special case when the two user classes merge into one class (i.e., a=b) and the number of users is infinite, the probability of blocking is known to be a hybrid of Erlang B and C formulas (Nesenbergs, 1979).

The probability of blocking for the common m-server facility of Figure 43 is treated in detail in Appendix B. Contrary to the concentration/switch server trunk problem of Section 5.2 and Appendix A, the present problem has the same blocking probability for both classes of users. A closed form solution for this probability is derived and discussed in Appendix B. We have computed a few exact cases thereof with a hand calculator. Several bounds and approximations are also indicated.

The situation is summarized in Figure 45. This figure shows the total load, La+Tb, that m servers are able to carry at 1% blocking probability. To get specific numbers, it is assumed in Figure 45 that L=50, T=10, and that a and b are such as to equate the line and trunk loads; i.e., La=Tb. Approximate curves are computed for the hybrid loss-delay parameter values $\theta=0$, 1/2 and 1. These curves fall roughly in the middle of a region that is defined by the so called bounds on m-server performance. The bounds are readily computed, as shown in Appendix B. However, they do exhibit a considerable region of uncertainty. For instance, when the total load is 12 Erlangs, the bounds indicate that anywhere from 10 to 20 servers may be required. The actual value turns out to be astonishingly near m=12. Note, by the way, that a loss system (with blocked calls cleared) has $\theta=0$ and requires fewer servers to provide the same grade of service as the $\theta>0$ systems.

5.4. Unresolved Traffic Issues for Switching Networks

The previous two sections have discussed two typical and useful traffic engineering problems. The resolution of such problems enhances the rationality of design and implementation. They have obvious effects on traffic efficiency, observed grade of service, the levels of user satisfaction and, last but not least, the costs of the military networks.

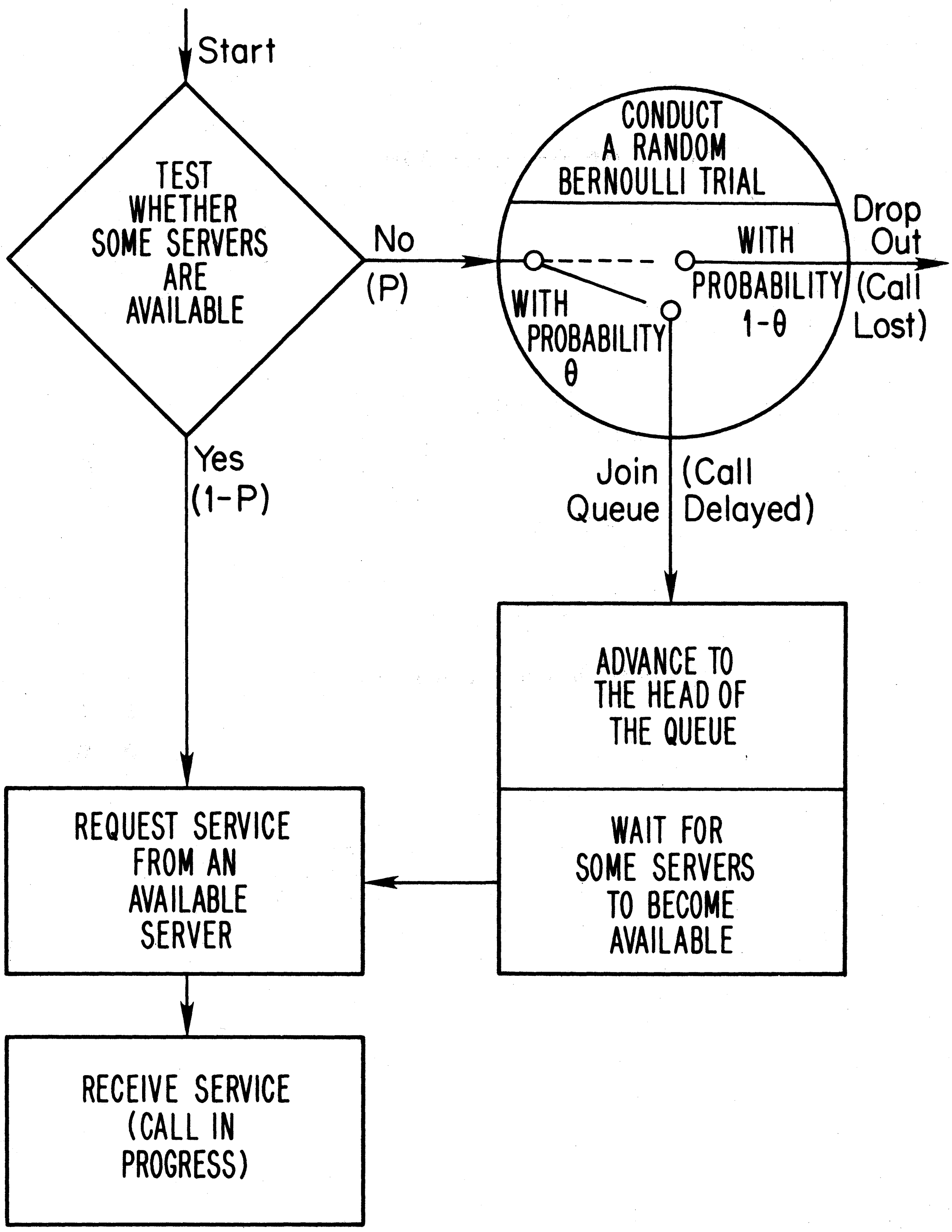


Figure 44. The traffic flow model for the loss-delay hybrid.

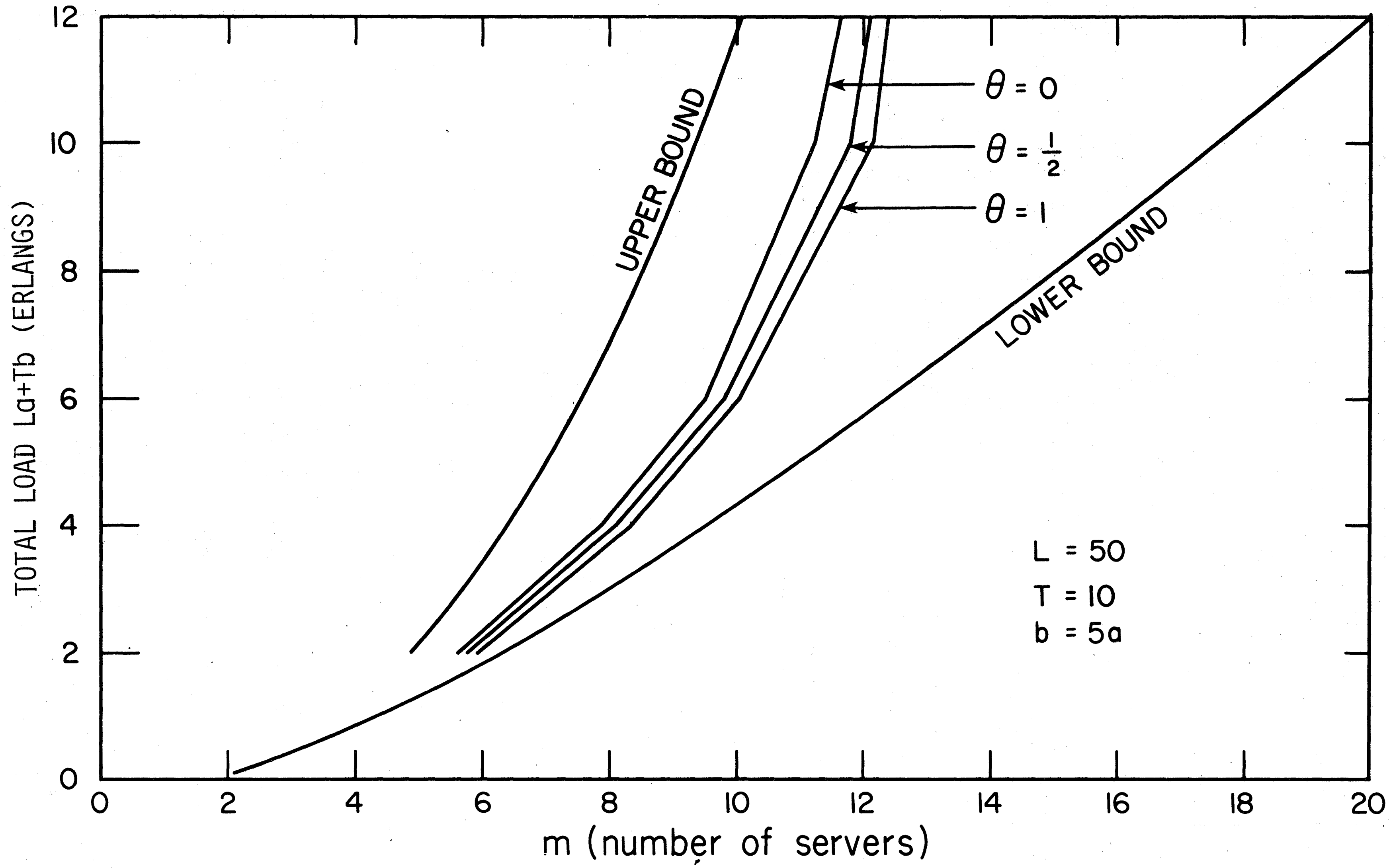


Figure 45. Bounds and estimated total load (La+Tb) versus required number of servers (m) for 1% blocking probability, parametric in hybrid parameter θ, and for the indicated values of L, T, a and b.

While necessary, the two earlier traffic engineering solutions (i.e., Sec. 5.2, 5.3 and Appendices A and B) are by no means comprehensive or all inclusive. They completely ignore the queueing time domain. That includes all waiting times, random delays, and the storage requirements associated with anticipated traffic-service processes. The area of queueing systems is, of course, broad and deep in both theory and applications (Saaty, 1961; Riordan, 1962; Kleinrock, 1975; Kleinrock, 1976b). It is also growing in impact on military voice and data telecommunications. For example, to complete a meaningful buffer size vs. line provision tradeoff, one has to grasp a number of tools. They include: network structuring, topological effects, capacity assignment options, traffic flow controls, and various alternatives for time sharing conflict resolution in multiaccess systems. The delay analysis, which ranges from estimation of average waiting times to somehow finding the 90, 95, or 99% occurrence levels for unknown distributions, is recognized as difficult at best. The analysis often appears untractable for real processes on real networks. Special chains of events, such as military actions, emergencies, peak overload failures, preemption, and even priority queueing, tend only to compound the analytical and engineering burdens. Simulation and observations on existing networks have been valuable (Kleinrock, 1976a; Kleinrock, 1976b).

There are also unresolved issues in the relatively better understood arena of grade of service. One such puzzling, but broadly applicable problem occurs when a network facility must handle at least three different kinds of traffic. The situation is illustrated in Figure 46.

Briefly, Figure 46 shows a switching network endowed with two types of input ports. On the left side there are L lines. On the right side there are T trunks. Service requests specify any of three distinct kinds of calls. As shown, the calls are denoted as "local," "distant," and "tandem." Local calls require two lines. Distant calls need one line and one trunk. Tandem calls use two trunks. The calls encounter a blocking event whenever they fail to find enough idle lines and/or trunks. For the three kinds of calls, there are three distinct blocking events with generally different probabilities. Blocked calls are presumed to be cleared or lost.

When the user population is finite, the problem appears to be a generalization of the rather involved problem discussed earlier in Section 5.2 and

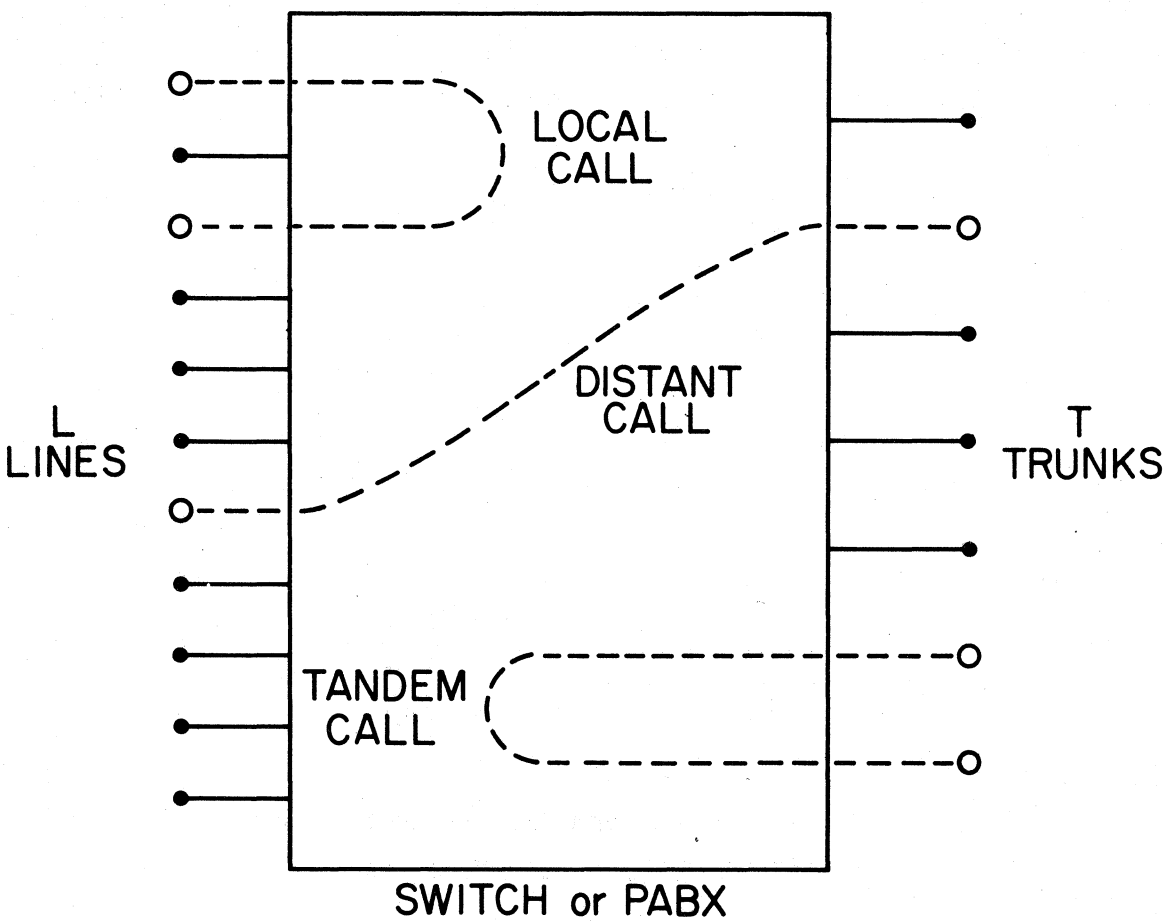


Figure 46. Three kinds of calls through a switching network.

Appendix A. To simplify, one assumes here an infinite user population; i.e., constant arrival rates for all three call types and under all conceivable system congestions. One further stipulates that the switch be nonblocking and with full availability. The call arrivals shall be generated by three independent Poisson sources, each with its own average rate. The average service or holding time shall be unity for all call categories.

Initial analysis indicates that this complex problem may nevertheless be tractable in a formal way. That is, closed form expressions may be feasible for state probabilities of interest. This includes those equilibrium states that constitute the three system blocking events. The computational effort involved in transforming the equations into useful numbers is at present time unexplored.

Appendix C outlines a preliminary approach to this problem, where, as shown in Figure 46, three groups of calls compete for the lines and trunks of a common service network.

6. APPLICATION TO FT. MONMOUTH ACCESS AREA

In the previous sections, the parameters which determine a switching system's capacity have been identified and some traffic-related performance parameters derived. In a specific switching environment, such as a military access area, these parameters can be used to specify the number, type and size of switching hubs needed to meet the non-tactical communication requirements for that area. One approach to sizing a digital access area switching system is demonstrated in this section.

It is important to define the term "access area." Access areas consist of military posts, camps and stations that share local communication facilities. The access area may include tactical units and other sub-elements whose planning and implementation has been the responsibility of a specific military department (MILDEP)[6] or the Tri-service Tactical Program Office (TRI-TAC).

[6] A recent Department of Defense Directive (Number 4605.5 dated February 7, 1979) provides consolidated unsecure telephone service to DoD components on a regional basis under a single management organization, the Defense Metropolitan Area Telephone System (MATS). Tactical elements are excluded. The area included in a MATS region has a 30-mile radius.

Currently, both tactical and non-tactical communications elements (i.e., terminals, switches and transmission facilities) in the access area are analog in nature. However, many of these elements are expected to transition into the digital world within the next decade. Local access area traffic that requires regional and interregional access must interface, at some point, with longer haul transmission facilities. These long-haul facilities may be commercial carriers or the global-strategic Defense Communications System (DCS). The worldwide DCS backbone network is controlled and operated by the Defense Communications Agency (DCA) for the combined military services.

Figure 47 illustrates a possible hierarchical network structure to provide access from a military base to the global backbone network. In this example, traffic from terminals is concentrated at a switch hub (PABX) via a star network. Several switch hubs are connected with a loop network to a central office on the base. Additional hubs may exist in nearby camps and stations in a local access area. More central hubs are located at other military bases in a larger region and one or more of them provides the interregional access via a long-haul backbone switch.

The hierarchical levels shown in Figure 47 can be related to the Ft. Monmouth, NJ, military bases as follows.

Switch Level	Access Level	Specific Area Examples
1	International	European Theater, Pacific
2	Interregional	Continental U.S.
3	Regional	Tri-service military bases in NJ
4	Local	Ft. Monmouth, NJ, Camp Woods, Camp Evans and CERCOM Building
5	Post, Camp, or Station	Ft. Monmouth, main post only
6	Office Complex	Squier Hall (CSA)

Regional and interregional access facilities for these specific areas are shown in Figures 48 and 49. Local and lower level access area digital switching and transmission requirements for Ft. Monmouth and its environs are considered in subsequent subsections.

Figure 48 covers the pmilitary region encompassing Ft. Monmouth and other military bases in the New Jersey area, including Army, Navy and Air Force facilities. Switching centers are interconnected by digital T-carrier.

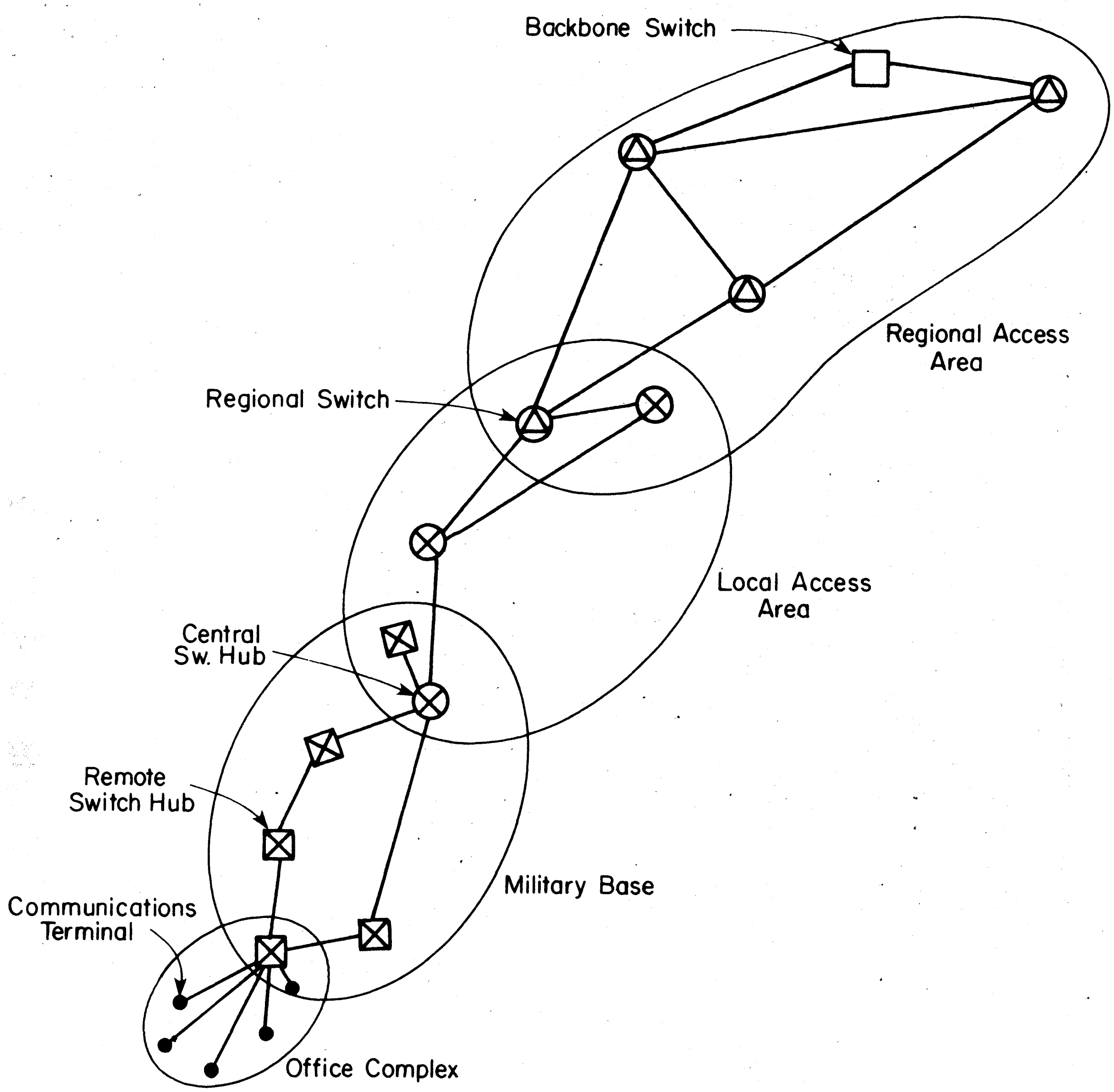


Figure 47. Hierarchical network configurations for interregional access.

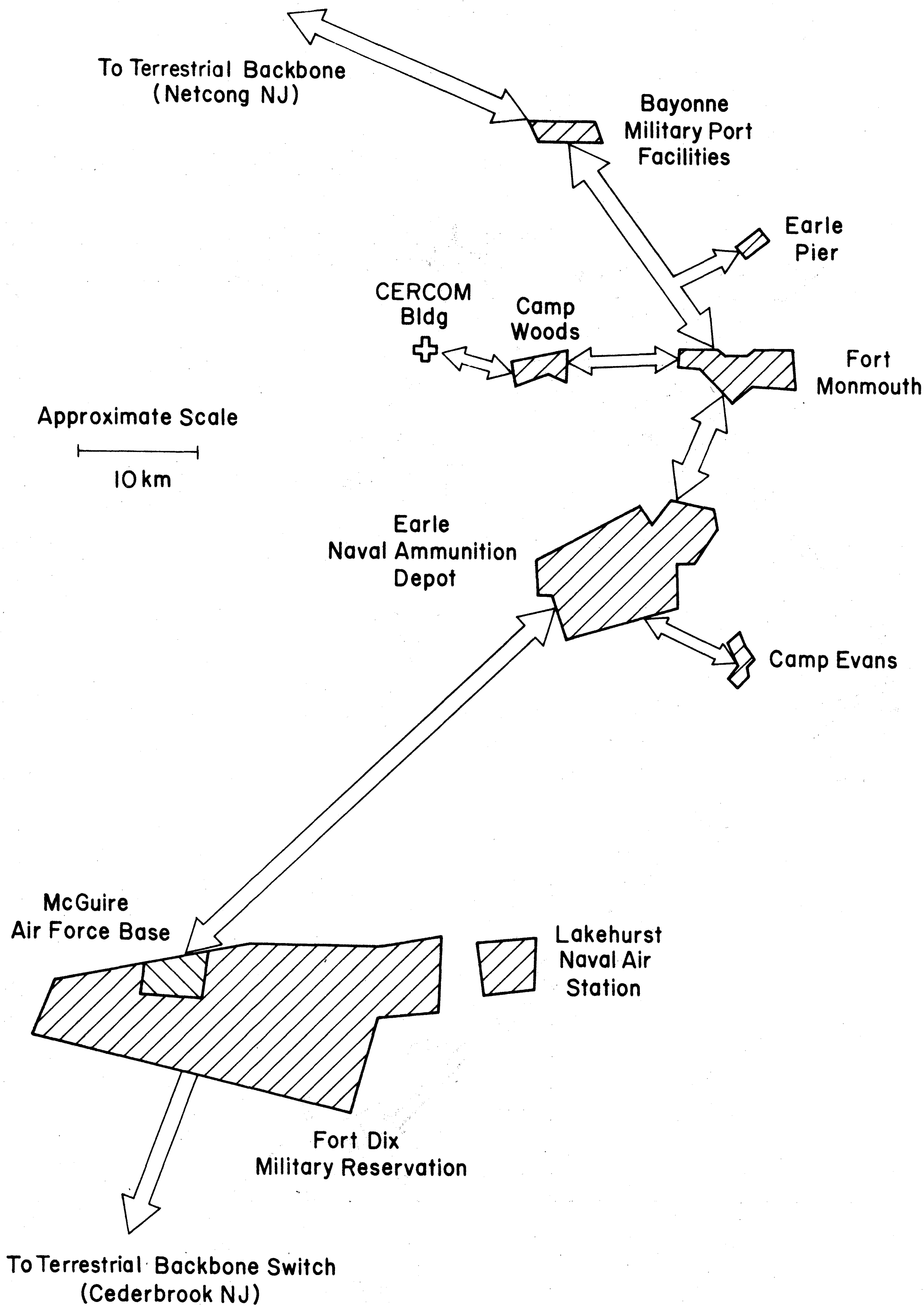


Figure 48. Regional access to the military reservations in New Jersey area via T-carrier.

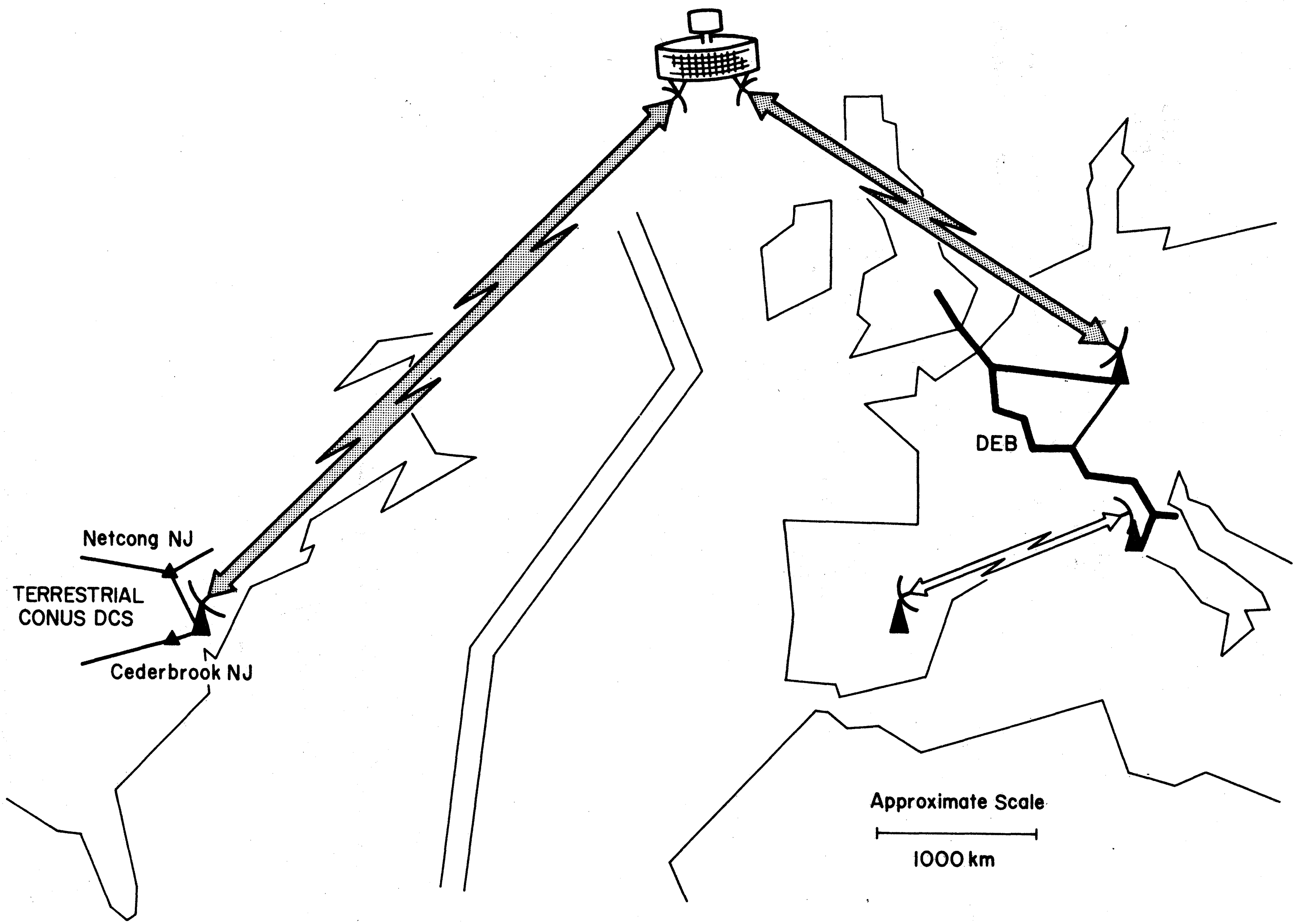


Figure 49. Strategic access to digital European backbone via DSCS.

Access to backbone switches is accomplished at both ends to enhance survivability.

Figure 49 depicts a possible strategic access link from this New Jersey region to the Digital European Backbone (DEB) via the Defense Satellite Communication System (DSCS), and hence to military bases in Western Europe. The DEB consists primarily of line-of-sight microwave links between backbone switches and repeaters. Troposcatter links permit access across less accessible regions to switching centers in Spain, Turkey, and Africa.

It has been projected that by 1980 there will be approximately 1500 military access areas worldwide. The number of communication terminals in these areas may range from small (50 to 300 terminals), to medium (300-2000) to large (over 2000 terminals) (see Wagner, 1977). Terminal densities in these areas may vary from less than 10 per square kilometer to 10,000 per square kilometer. Table 15 gives estimates of the number of communication terminals of various types located on a medium size post, a relatively large access area, a military region, the continental United States (CONUS), and including overseas (OCONUS).

Additional pertinent parameters for planning future non-tactical digital communications networks are listed in Table 16.

In the following subsections, the key parameters listed in the table are quantified for the Ft. Monmouth access area. The area includes the main post, adjacent camps and other nearby Army installations. Switching requirements are determined using the most recent parameters. Future growth is not considered. Several assumptions are made to simplify this planning approach. They are given in the appropriate sections.

One must emphasize that the purpose here is to demonstrate the planning concept, to indicate how switch hub requirements can be determined, and to show how switch capacities may be specified. The results obtained, therefore, should not be construed as a final engineering design. Additional detailed analysis is required to determine whether subsystems satisfy the performance objectives and whether final system capacities can meet projected growth and traffic patterns for the area. Digital terminals and data circuits are not included in this simplified planning approach. The advent of computer processing, intelligent terminals and automated business equipment is expected

Table 15. Estimates of Military Communications Terminals, Local to Worldwide

	Medium Post	Access Area	Military Region	CONUS ---	OCONUS ---
Area (km^2)	5	50	5,000	15M	500M (Worldwide)
Population (Military & civilian)	5,000	15,000	50,000	1.5M	0.5M
Telephones					
Mainline	1,500	6,000	15,000	300,000	50,000
Extensions	1,000	4,000	10,000	200,000	30,000
Sub-totals	2,500	10,000	25,000	500,000	80,000
Data Terminals					
Interactive	10	30	100	1,000	400
Computer	2	6	30	200	50
Narrative	5	15	50	500	100
Facsimile	3	8	25	250	50
Data	2	6	30	300	50
Sub-totals	22	65	235	2,250	650
Secure Terminals					
AUTOSEVOCOM	1	3	10	1,000	500

Table 16. Parameters Required for Access Area Planning

Access Area Profile

- o Geographic Area
- o Population
- o Terminal Types
- o Terminal Densities
- o Traffic Statistics

Expected Growth Patterns

- o Mainline Telephones (basic and enhanced service features)
- o Data Terminals and Circuits

Service Features

- o Basic
- o Enhanced

Service Classes

- o Normal
- o Restricted
- o Operator Assisted
- o Maintenance
- o Special (digital secure voice terminal)

Performance Objectives (user oriented)

- o Grade of Service
- o Acceptable Delays
- o MTTR, MTBF

Trunking Requirements

- o Interregional
- o International
  - Commercial
  - WATS
  - Specialized Common Carrier
  - AUTOVON, AUTODIN and AUTOSEVOCOM

Table 16 (cont.)

Access Nodes

- o Number and Location of Remote Concentration Hubs
- o Number and Location of Remote Switching Hubs
- o Number and Location of Central Switching Hubs

Switch Parameters

- o Subscriber's Lines Served (opening)
- o Subscriber's Lines Served (ultimate capacity)
- o Line Distribution for Various Classes of Service
- o Interswitch Circuits, Initial and Final Capacity
- o Busy Hour Call Attempts
- o Peak Traffic Margins
- o Service Features (Basic and Enhanced)
- o Reserve Line, Trunk and Processing Capacity for Traffic Growth

Interswitch Circuits

- o Initial and Ultimate Capacity
- o Links to Remote Hub and Size of Groups
- o Links to Tandem Hubs and Size of Groups
- o Special Lines (emergency, maintenance, information)
- o Mode of Operation (incoming, outgoing, both ways)
- o Control Signaling Technique
- o Type of Transmission Facilities
- o Alternate Routing

Administration and Maintenance

- o Accounting
- o Fault Monitoring
- o Numbering Plan
- o Alternate Routing

Table 16 (cont.)

<u>Measures of Effectiveness</u> o Reliability o Availability o Survivability o Cost o Security o System Life Span

to have considerable impact in the future on access area digital switching systems, the network transmission facilities, and the users they serve.

6.1. Access Area Profile

The Ft. Monmouth access area includes the main post, Camp Woods, the CERCOM Building, Tri-Tac offices and other nearby installations as shown in Figure 50. Ft. Monmouth is the headquarters of the U.S. Army Communications and Electronics Readiness Command (CERCOM) and the Communications Research and Development Command (CORADCOM). Other tenants include the Communication Systems Agency (CSA), the Satellite Communications Agency, elements of the Electronics Research and Development Command, and the Aviation Research and Development Command (AVRADCOM). The TRI-TAC program office and the JINTACCS (Joint Integrated Tactical Command and Control Systems) program office are also nearby.

A map of the main post showing major office complexes, apartment and barrack facilities, and other pertinent installations is shown in Figure 51.

The Ft. Monmouth access area profile information was obtained in an earlier study by Nesenbergs and Linfield (1976). The demographic results are summarized in Table 17; the telephone environment in Table 18; and the data circuit inventory and computer facilities in Table 19. Table 20 estimates the number of terminals by type, to be used in the area around 1985. If this projected growth should occur (doubling the total number of voice and data terminals), it would place major demands on the number and capacities of the switching systems needed.

Using Figure 51 and Table 17, the main post can be divided into subareas which show the population distribution as a function of density. This is done in Figure 52. Table 21 indicates the density (terminals or drops per square kilometer) used for each type of area. The table also shows the total subarea in km^2 of each type for the major posts, camps and office buildings. This density profile is useful for optimizing network configurations and for arriving at gross cost estimates for the distribution system (Nesenbergs and Linfield, 1976).

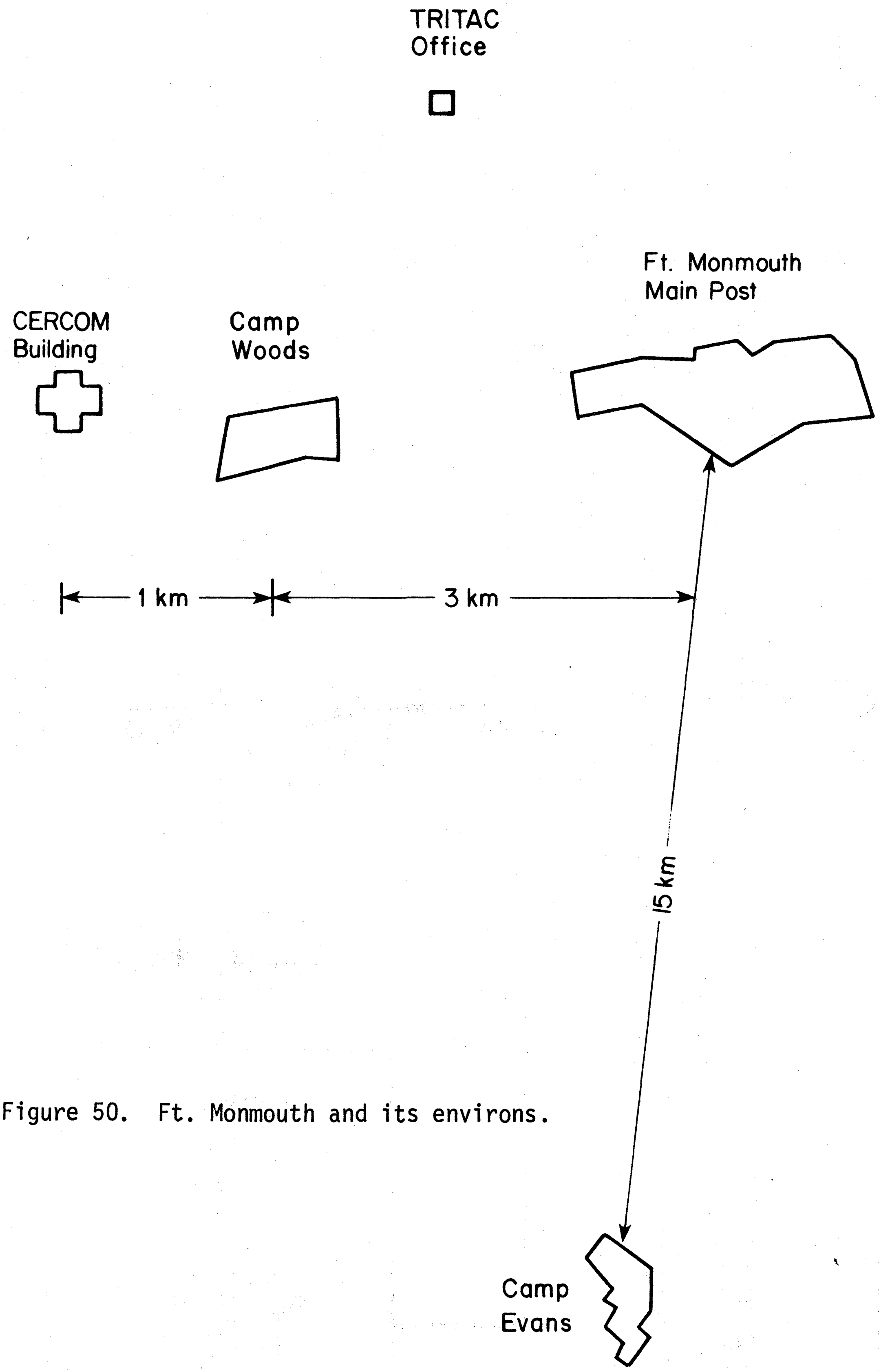


Figure 50. Ft. Monmouth and its environs.

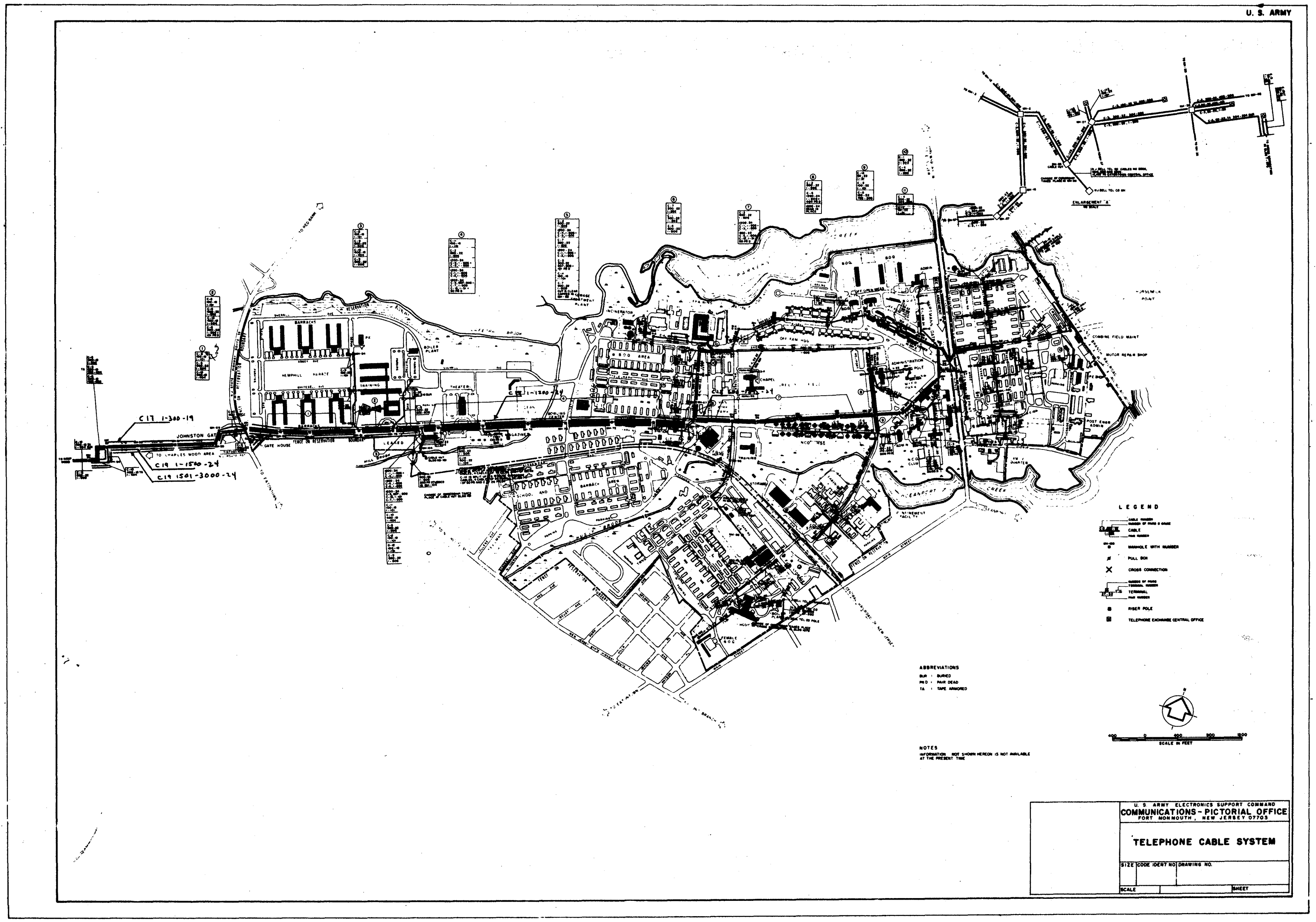


Figure 51. Ft. Monmouth main post.

Table 17. Ft. Monmouth Access Area Demographics

Population by Subarea

Subarea	Main Post	Camp Woods	CERCOM Building	Camp Evans and Others	Total Population
Office Complex	1,284	2,205	3,800	383	7,672
Apartments	1,830	150	--	422	2,402
Service	356	40	--	--	396
Barracks	200	--	--	--	200
High Residential	480	4,136	--	--	4,616
Low Residential	--	--	--	8	8
	4,150	6,531	3,800	813	15,294

Population By Major Office Complex

Complex	Location	Population
Vail Hall, DCO	Main Post	90
Patterson Hospital	Main Post	500
Squier Hall, CSA	Main Post	350
SATCOM Office	Main Post	240
Administration Bldg.	Main Post	104
Hexagon	Camp Woods	2000
Administration	Camp Woods	260
EW Laboratory	Camp Woods	145
TriTac Office	New Shrewsbury	180

Table 18. Ft. Monmouth Access Area Telephone Environment

Station Apparatus By Subarea

Types	Post ARTADS TriTac	Camp Woods	CERCOM Building	Camp Evans	Totals
Telephones					
Mainline	1,775	1,455	1,625	457	5,312
Bridged	2,064	1,059	1,614	401	5,138
Secure	1	1	1		3
Totals	3,840	2,515	3,240	858	10,453

Dial Central Offices

	Main Post	Camp Woods	CERCOM Building	Camp Evans	Totals
Designator	DCO-1 (Vail Hall)	DCO-2 (Hexagon)	--	DCO-3	3
Total Capacity	4,800	2,200	--	700	7,700
Working Capacity	--	--	--	--	5,312

Access Area Trunking

Internal	351
Special Toll	7
Commercial Incoming	102
Commercial Outgoing	70
WATS Lines	21
AUTOVON	94
Total Trunks	645

Table 19. Ft. Monmouth Access Area Data Terminal Environment

Data Circuit Inventory

Circuit Bit Rate	Main Post	Camp Woods	CERCOM Building	Camp Evans	Totals
50 kb/s	2	-	-	-	2
19.2 kb/s	2	-	-	-	2
7.2 kb/s	2	-	-	-	2
4.8 kb/s	7	-	1	-	8
2.4 kb/s	3	1	-	-	4
2.0 kb/s	-	-	1	-	1
1.2 kb/s	1	-	-	-	1
900 b/s	-	-	4	-	4
134.5 b/s	2	-	-	-	2
100	2	20	2	10	34
Test	1	10	6	-	17
Totals	22	31	14	10	77

Large Computer Facilities

	Computer Types
Northeast Computer Center	1 - IBM 360/50
Bldg. 1152, Main Post	1 - IBM 360/65
Hexagon, Camp Woods	2 - Burroughs 5500
CERCOM Building	1 - IBM 360/65

External Data Services

Dedicated Lines	18
Dialup Lines	30
Total	48

Table 20. Number and Types of Terminals Projected for Ft. Monmouth Access Area in 1985

Telephones (Mainline)	
o Analog (4 kHz)	8,000
o Digital (64 kb/s)	1,000
o Video (6 Mb/s)	5
Computer Access	
o Low Speed (450 b/s)	50
o Medium Speed (3.6 kb/s)	20
o High Speed (48 kb/s)	10
Teletype	
o Low Speed (300 b/s)	100
o High Speed (4.8 kb/s)	20
Facsimile	
o Low Speed (9.6 kb/s)	10
o High Speed (56 kb/s)	5
o Color (1.5 Mb/s)	5
Television	
o CATV (1-way Channels)	8
o Common Channel (2-way Channels)	4
o Slow Scan (2-way Channels)	6

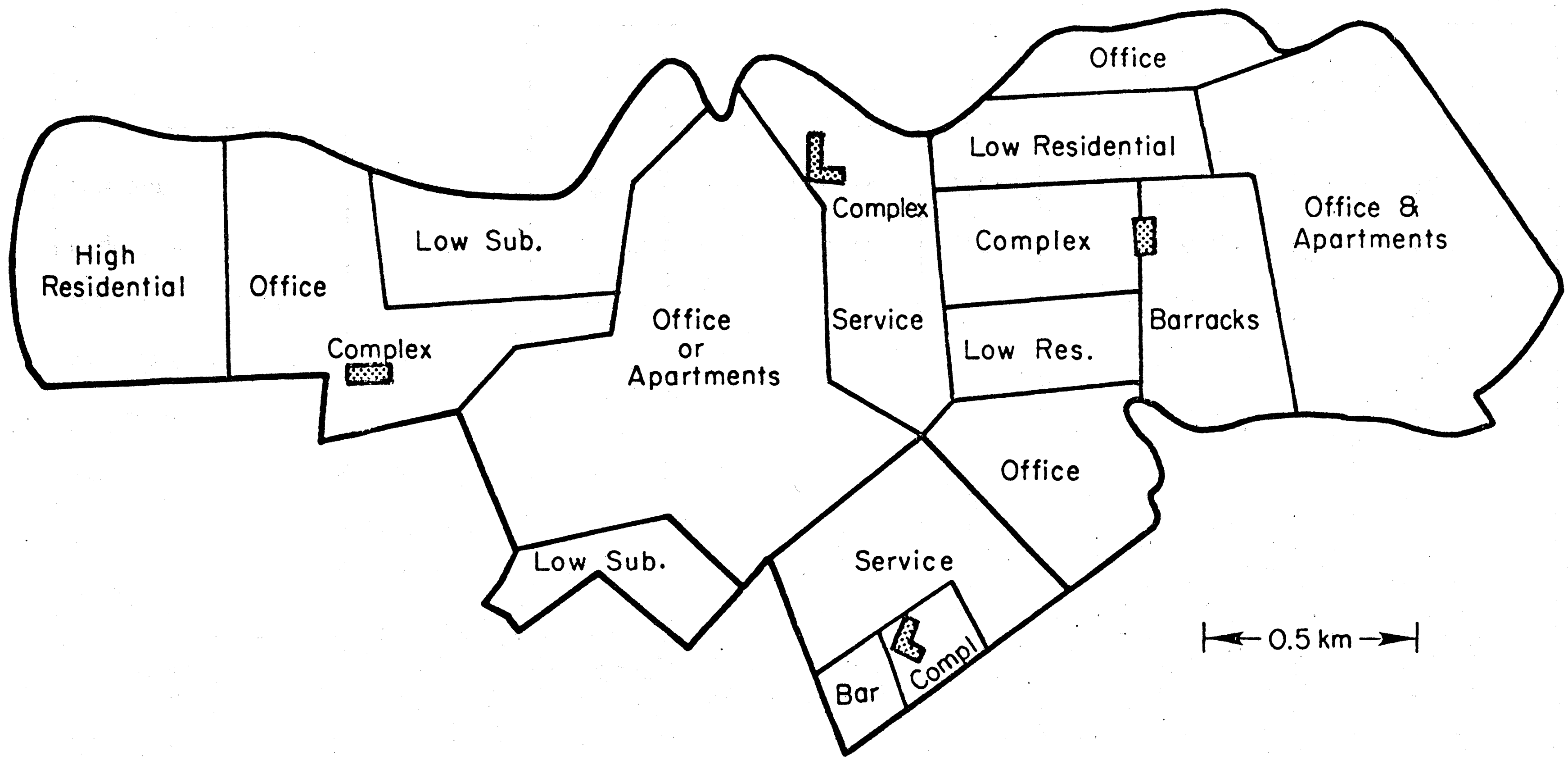


Figure 52. Population distribution for Ft. Monmouth.

Table 21. Terminal Density Profile

Area Type	Density Drops/km^2	Subarea Estimate (km^2)				Total Area (km^2)	Total No. of Drops
		Post	Woods	CERCOM	Evans		
Complex	10,000	0.15	0.15	0.33	0.05	0.68	6,800
Office	5,000	0.20	0.15	----	0.05	0.40	2,000
Service	2,000	0.40	----	----	0.02	0.42	820
Barracks	1,000	0.15	----	----	----	0.15	150
High Residential	500	0.50	0.60	----	----	1.10	550
Low Residential	200	0.30	0.10	----	0.05	0.45	90
High Sub.	100	0.20	0.10	----	0.20	0.50	50
Low Sub.	50	0.10	----	----	0.10	0.20	10
Totals		2.00	1.10	0.33	0.47	3.90	10,470

6.2. Basic Assumptions and Implementation Concepts

The implementation concept for an all-digital access area network is illustrated in Figure 53. Digital and analog terminals interface with the network via concentration units. A time multiplexed loop connects these units to the switching hub. Repeatered lines may be used for units located more than 2 kilometers from a digital switching hub. Trunk circuits are also shown as time division multiplex loops. They can be multiplexed to higher levels to provide time slots for 24 (T1), 48 (T1C) and 96 (T2) trunks. Other details and assumptions made for each subsystem of the network are given in the following paragraphs.

Remote Switching Hubs (RSH)

Remote switching hubs, like private automatic branch exchanges (PABX), are located in office complexes and other areas of high terminal density. All input and outputs from the switching matrix contain 24 time slots. Lines interface via local or remote concentrators. Intra-area trunks interface directly (24 channels) or via multiplexers (48 and 96 channels). Some of the remote switch functions such as fault monitoring, maintenance and diagnostics, accounting, and other special services are controlled from a central switch hub using a common-channel signaling circuit. However, the remote switch can serve its subscribers independently, with some reduction in service features offered.

Switching is performed by the matrix using time slot interchanges on each loop and time multiplexed gates between the loops, as depicted in Figure 54. The switch size, in terms of number of terminations, depends on the number of TDM loops used and on the line concentration ratio. The switching matrix is assumed to be non-blocking with full availability to all lines and trunks.

Local and Remote Concentration Hubs (LCH and RCH)

Concentration hubs may be colocated with a switch hub or located remotely in areas of high terminal density. All typical mainline circuits terminate on either an RCH or an LCH. The outputs are 24-channel time-division multiplexed loops. Up to 40 lines may terminate on each hub. This provides a blocking probability of 0.01 for calls to other hubs and 0.02 for local calls. This assumes that 25% of the originating calls are distant calls and

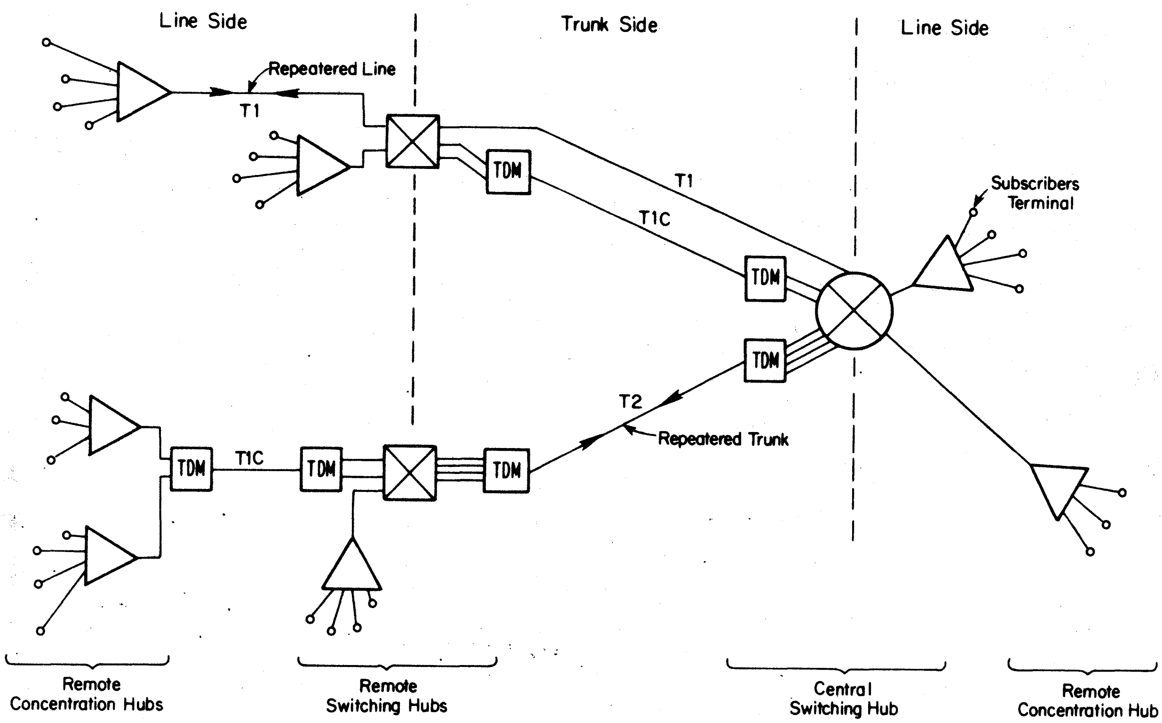
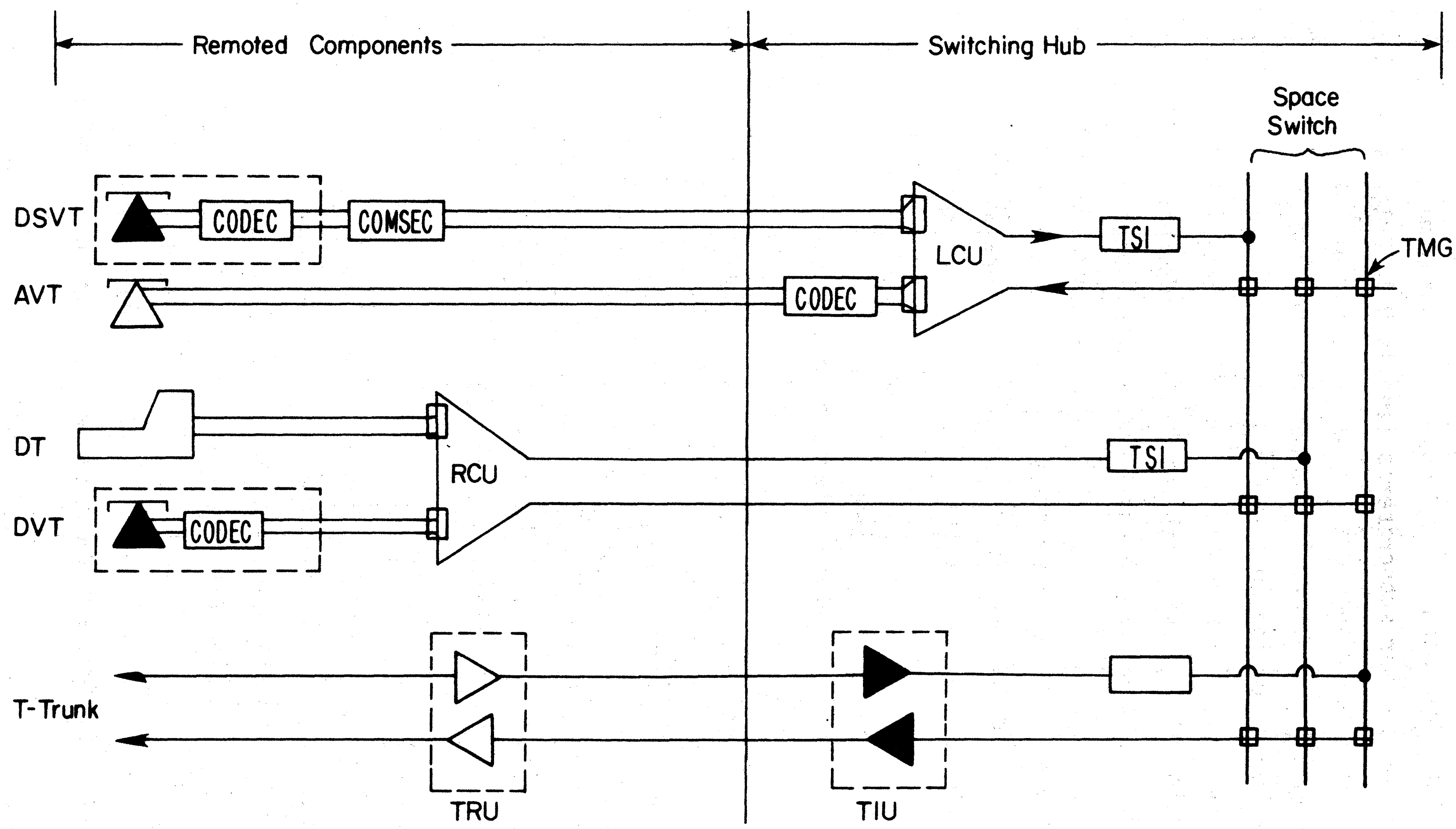


Figure 53. Implementation concept for digital concentration, multiplexing, switching and transmission in the access area.



Legend

DSVT:	Digital Secure Voice Terminal	RCU:	Remote Concentration Unit
AVT:	Analog Voice Terminal	TRU:	Trunk Repeater Unit
DT:	Digital Terminal	TIU:	Trunk Interface Unit
DVT:	Digital Voice Terminal	TSI:	Time Slot Interchange
LCU:	Local Concentration Unit	TMG:	Time Multiplexed Gate

Figure 54. Local and remoted components of a switching hub.

75% are local calls; i.e., a/b=1/3. Figure 41, in Section 5, indicates that under these conditions the 24-channel loop can handle approximately 5 Erlangs. The average traffic intensity per subscriber's line is 0.12 Erlangs (4.3 ccs).

The traffic intensity in Erlangs which flows through the switching matrix is derived using Figure 55. It is assumed that, on the average 10% of the calls homing on a concentration unit are trunk calls, 15% are calls to another concentration unit on the same switch, and 37.5% are calling another 37.5% on the same concentrator. This high community of interest at the concentrator level could be expected in an office complex where many inter-office calls are typical. The switching matrix load generated by a single multiplexed loop is the sum of the three types of call; i.e., 0.5+1.87+0.75= 3.12 Erlangs. It is assumed that all concentration units are similar, so that for k such multiplex loops the total switching matrix capacity is k(3.12) Erlangs.

Terminals

The majority of the terminals which interface on a concentration hub are analog telephones. Both 2-wire and 4-wire lines consisting of twisted wire pairs are used. Station apparatus may provide basic or enhanced service features depending on the apparatus used and the interface. Analog-to-digital conversion is normally provided by the interface codec. This codec can be remoted with 4-wire line to the telephone set and communication security provided with the addition of a COMSEC unit. Data terminals can also be served using special interface units. See Figure 54. In Section 6.3 it is assumed that only analog voice terminals with basic service features are used.

Central Switching Hub (CSH)

The central switching hub provides all of the functions and features of a remote switching hub. In addition, the CSH is the master control for the intra-area network and for interswitch trunking. It performs all the local administrative functions and provides the access to the public switched network, specialized common carrier networks, the DCS backbone, and gateways to other networks. Transmission facilities to these external networks may be analog or digital. The mode, code and speed conversions are performed by central switch elements.

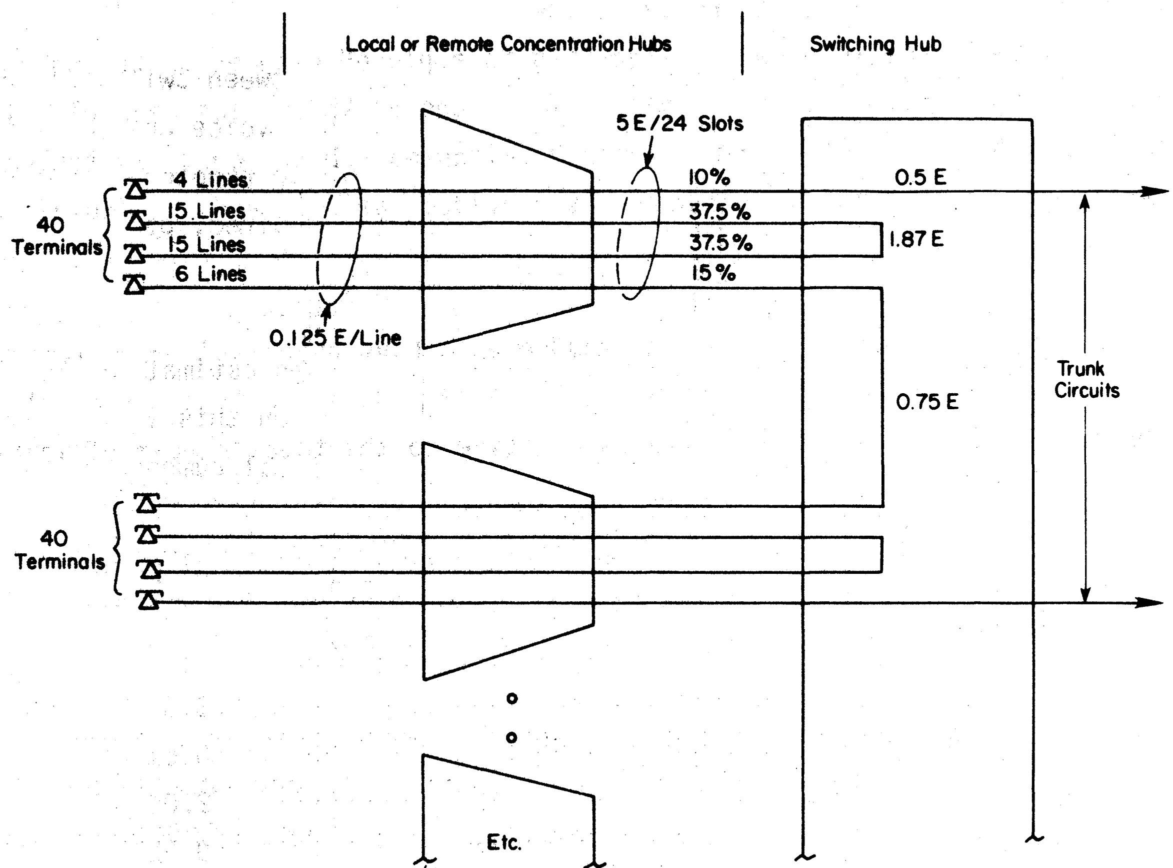


Figure 55. Traffic flow assumed for remote switching hubs.

Control Signaling

It is assumed that the majority of the control signaling for line supervision, addressing, alerting, etc. is accomplished on a per-channel basis. On digitized voice circuits that use PCM, the least significant bit is borrowed from each channel in every sixth frame for signaling purposes. For data circuits, an 8 kHz signaling channel is used when digitized voice or other channels of that rate are available.

Common-channel interswitch signaling is employed between switching hubs. That common channel is usually equivalent to one 64 kb/s voice channel. This provides remote control from the central switching hub to remote switching hub. Signaling to external networks is provided, as required, by individual networks.

Trunking

Intra- and interarea trunking requirements have been estimated from the current trunking requirements given in Table 18. Based on this table, the required percentage of trunk circuits relative to the total number of main-line terminations is as follows.

Remote to Central Switching Hubs	10.0%	
Intra-area Switch Hubs	7.0%	
Interregional Switching	6.0%	
o Commercial		3.0%
o WATS Line		0.5%
o DCS Backbone		2.0%
o Other		0.5%

It is further assumed that all intra- and interregional trunks are multiples of 24-channel digital carrier, thus the above percentages can only be approximated to the nearest upper T-carrier defined in Table 22. The traffic intensity values for blocking probabilities, P=0.01 and P=0.001, are based on infinite sources, full availability, and blocked calls cleared; i.e., on the Erlang B equation. In Section 6.3, it is assumed that the intra-area trunks have a blocking probability of 0.001.

Table 22. Intra- and Interregional Trunking Facilities

Carrier Designator	Number of Voice Channels	Erlangs Carried P=0.01	Erlangs Carried P=0.001	Signaling Channels
T1	23	14.5	11.5	1
T1C	47	35.2	30.1	1
T2	95	79.4	70.9	1

Figure 56 illustrates the trunking requirements derived for the Ft. Monmouth access area. Time division multiplexers for high level carriers are assumed, but not shown. Repeatered lines are indicated for path lengths exceeding 2 kilometers. Access to the DCS backbone is shown as a T2 (96 channel) circuit to Cederbrook, NJ. An additional T1 or T2 circuit could be added to Netcong, NJ, for redundancy.

6.3. Digital Switching Requirements for Main Post and Nearby Installations

The implementation concept described in the previous section can be applied to the main post at Ft. Monmouth using the post profile data of Section 6.1.

A detailed map of the Post was shown in Figure 51. Using an overlay on a larger version of this map and the information in Tables 17, 18, 19, and 21, one identifies approximate terminal cluster locations for remote concentration hubs and remote switching hubs. This is done in Figure 57. The central switching hub is located in Vail Hall. It provides the intra-area and interregional access trunks. The switch also performs local switching and administrative functions and, when necessary, tandem switching between other central hubs.

Figure 58 is a composite diagram of the total switching requirements for the access area with the geographic aspects removed. The switching and termination requirements are depicted for four hierarchical levels, intra-switch, intra-post, intra-area, and interregional. Only the mainline telephone terminations are shown on the figure. They total to just over 5300 mainline telephones for the entire access area. The central switching hub located at Vail Hall provides access to telephones in Vail Hall and its immediate surrounding area. It also provides tandem switching.

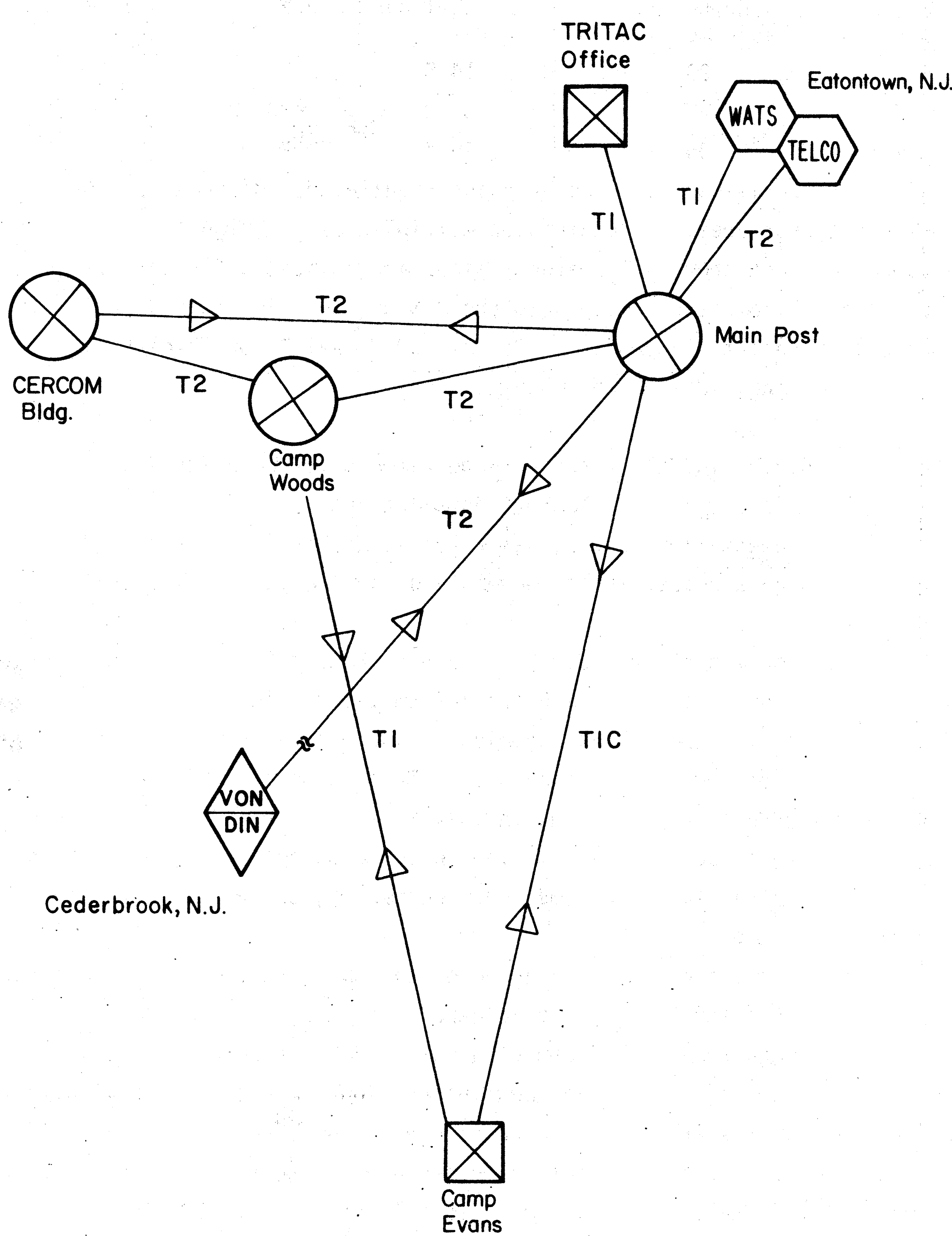


Figure 56. Trunking from main post to environs, Ft. Monmouth, NJ.

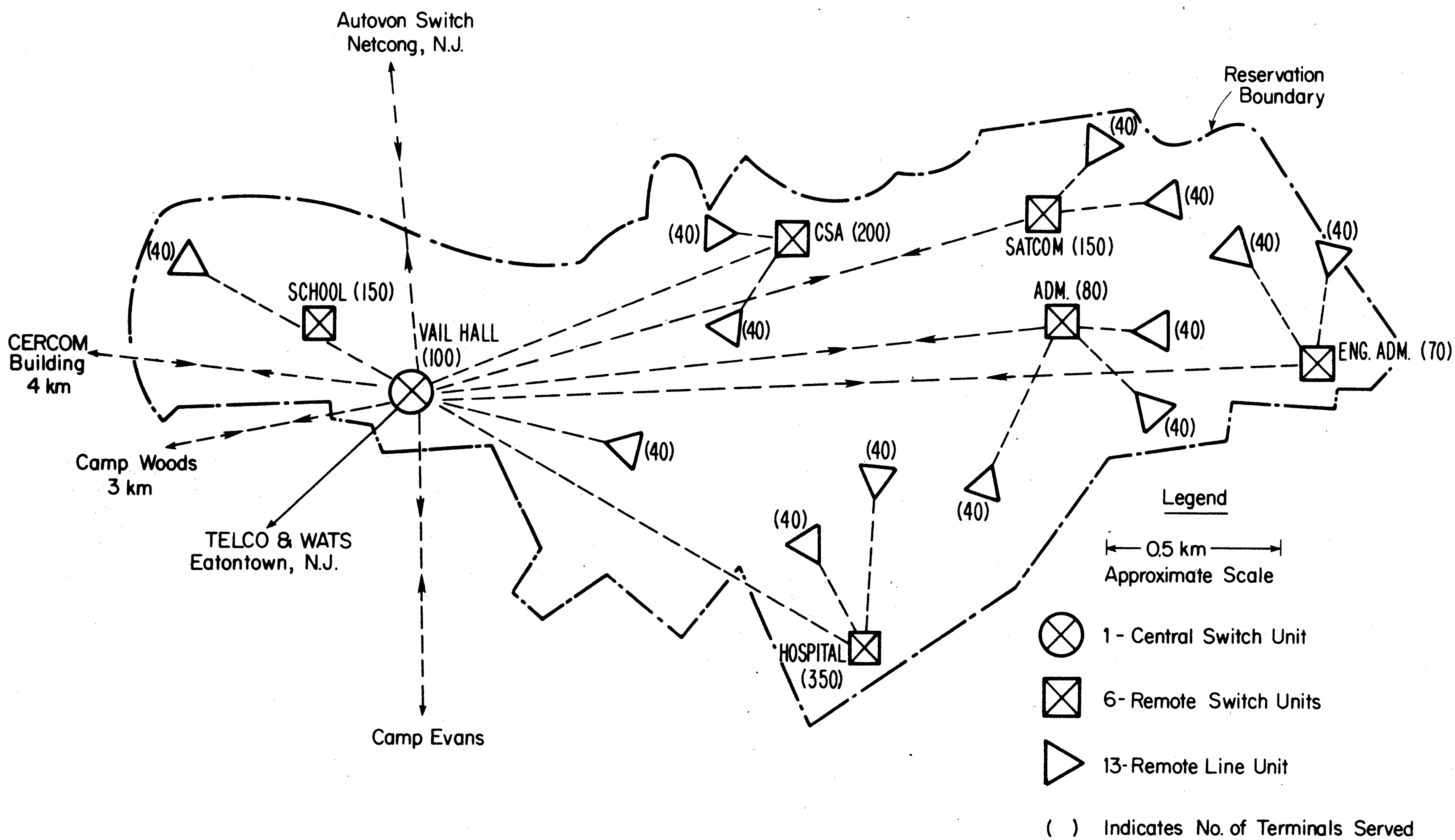


Figure 57. Concentration, switching, and trunking on main post, Ft. Monmouth, NJ.

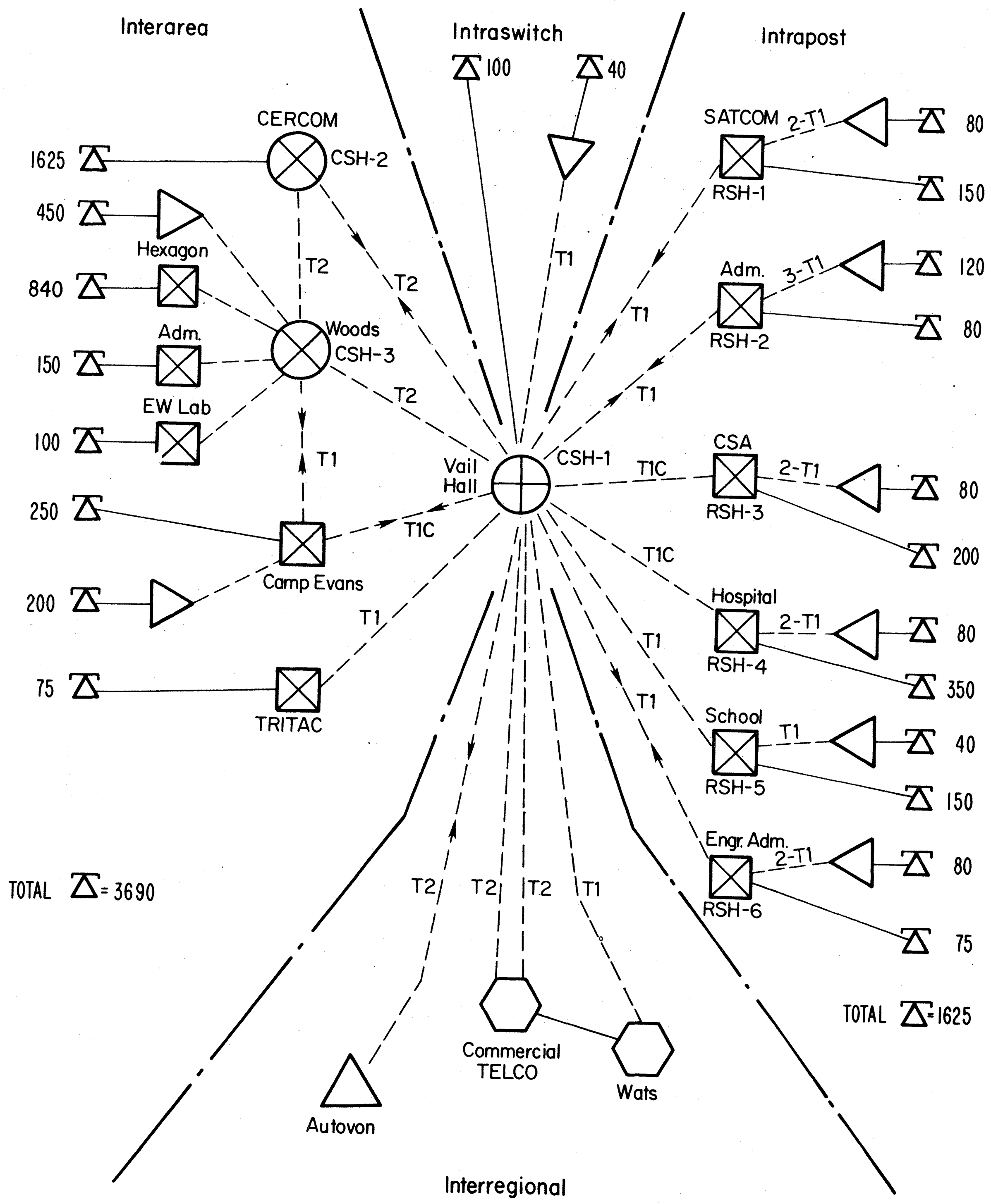


Figure 58. Digital concentration, switching and trunking requirements for mainline telephones in Ft. Monmouth access area.

The mainline terminations, trunk circuits, matrix capacity and processor capacity needs for the six remote switching hubs (RSH) on the main post have been estimated. The projected requirements are listed in Table 23. This table also indicates the number of remote concentration units and local concentration units required for each hub. It assumes that four trunk channels are required for each concentrator (i.e., 10% of the concentrator's input terminations). The table assumes that each 24-channel concentrator output loop carries 5 Erlangs of traffic and that the switching matrix handles 3.12 Erlangs per concentrator loop. Finally, the processor capacity is estimated to be approximately 20 times the matrix capacity, based on 3 minute holding times.

These remote switching hub characteristics do not include reserve capacity for traffic growth, nor do they provide processor capacity margins for traffic peaks. Only the basic features are assumed offered, but to all mainline telephones. The processor capacity would have to be increased as more full features sets are added.

The factors which constitute the capacity requirements of the main post central switching hub are summarized in Table 24. This hub serves subscriber terminals located in Vail Hall and its immediate area. Its main function, however, is to serve as a tandem trunking switch for remote switching hubs on the main post, and for other central offices and hubs in the local access area; e.g., the CERCOM Building, Camp Woods, Camp Evans and TRI-TAC offices. At the same time, it provides all switching to inter-regional trunks for the local access area; e.g., AUTOVON, commercial telephone company, and for wide-area telephone service (WATS). Access to the Federal Telephone System (FTS) is not included. Incoming FTS calls to the area are connected via commercial circuits.

Table 24 lists the number of terminals served and the number of trunk channels required from each location. These projected trunking requirements are approximately the same as those in use now (see Sec. 6.2). Given the number of trunk channels required, a T-carrier is selected which provides a sufficient closest multiple of 24 channels. The capacity per carrier is obtained from Table 22 for an assumed trunk blocking probability of 0.1%.

The total load carried by intra-post, inter-area and interregional trunks is 536.6 Erlangs. Since these trunks handle both originating and

Table 23. Main Post Remote Switch Hub Capacity Requirements

Hub Designator	Hub Location	RCH's	LCH's	Mainline Terminations	Conc Loops (T1)	Trunk Channel[1]	Conc Traffic[2] (Erlangs)	Matrix Capacity[3] (Erlangs)	Processor Capacity[4] (BHCA)
RSH-1	SATCOM	2	4	230	6	24	30	18.72	375
RSH-2	Adm. Bldg.	3	2	200	5	20	25	15.60	312
RSH-3	Squier Hall	2	5	280	7	28	35	21.80	436
RSH-4	Hospital	2	9	430	11	44	55	34.30	686
RSH-5	School	1	4	190	5	20	25	15.60	312
RSH-6	Engr. Adm.	2	2	155	4	16	20	12.50	250

Notes:

(1) 4-trunk Channels per Concentrator Loop

(2) 5 Erlangs per Loop

(3) 3.12xNumber of Loops (see Fig. 55)

(4) 20xMatrix Capacity for 3-minute Holding Times

Table 24. Main Post Central Switching Hub Line and Trunk Loading

	Source Designator	Location	Terminals Served	Trunk Channels	Loops or T-Carriers Required	Carried Load (Erlangs)
Lines	Intraswitch					
	RCH	Vail Hall Area	40	4	1	3.1
	LCH	Vail Hall	100	10	3	9.4
	Line Total		140	14	4	12.5
Trunks	Intrapost			(~10%)		
	RSH-1	SATCOM	230	24	T1	11.5
	RSH-2	Adm. Bldg.	200	20	T1	11.5
	RSH-3	Squier Hall	280	28	T1C	30.1
	RSH-4	Hospital	430	44	T1C	30.1
	RSH-5	School	190	20	T1	11.5
	RSH-6	Engr. Adm.	155	16	T1	11.5
	Subtotal		1,485	152	8-T1 (equiv.)	106.2
	Interarea			(~7%)		
		CERCOM	1,625	114	T2+T1	70.9
		Camp Woods	1,540	108	T2+T1	82.4
		Camp Evans	450	32	T1C	30.1
		TriTac Office	75	5	T1	11.5
	Subtotal		3,690	259	13-T1 (equiv.)	194.9
	Interregional			(~6%)		
	AUTOVON	Netcong, NJ	2% of 5,315	103	T2	70.9
	Commercial	Eatontown, NJ	3% of 5,315	155	2-T2	141.6
	WATS Lines	Eatontown, NJ	0.5% of 5,315	25	T1	11.5
	Other		0.5% of 5,315	25	T1	11.5
	Subtotal			308	14-T1(equiv.)	235.5
	Total		5,315	719	34-T1(equiv.)	536.6

terminating calls, the load carried by the matrix for these trunk calls is 536.6÷2=268.3 Erlangs. Adding the 12.5 Erlangs for intraswitch line calls gives 280.8 Erlangs for the total matrix capacity. If one assumes an average holding time per call of 3 minutes, then the processor must handle 20x280.8=5616 call attempts per hour on the average.

These results, obtained from Table 24, are summarized in Table 25 for the CSH. The results from Table 23 for RSH-3 located in Squier Hall are also summarized in Table 25. Both required and engineered capacities are given in Table 25. The engineered capacity includes a less than or equal to 20% increase in matrix load and processor capability to handle peak traffic loads. No margin is reserved for traffic growth in this example.

A switch system's capacity can be estimated as follows. Assume a 3-minute holding time per call. Then approximately (Sec. 4.5),

$$\text{Processor Capacity (BHCA)} = 10\ A_T M,$$

$$\text{Matrix Capacity (Erlangs)} = 1/2\ A_T M.$$

where A_T is the average traffic intensity (or load) per termination and M is the total number of terminations. These relationships are plotted in Figure 59. The capacity results from Table 25 for the central (CSH-1) and remote (RSH-3) switching hubs and some commercial and military switch capacity estimates are also indicated on this figure.

The CSH-1 is primarily a tandem trunk switch. The average intensity per trunk termination is relatively high compared to the line intensity of terminals homing on RSH-3. The latter, of course, is primarily a line switch load. The difference is typical for these two switch types. For a switch with a more equal number of lines and trunks, the capacity requirement can be estimated by assuming independence and by adding the results obtained for each type of termination. For example, assume a switch with 200 line termination with 0.2 Erlangs/line, and 100 trunk terminations with 0.5 Erlangs/trunk. Then, see Figure 59, the following requirements are obtained for the system.

Required Capacity	Lines	Trunks	Totals
Terminations (Total No.)	200	100	300
Processor (BHCA)	200	500	700
Matrix (Erlangs)	10	25	35

Table 25. Summary Characteristics for Central (Main Post) and Remote (Squier Hall) Switching Hubs

CSH-1, Main Post	Required Capacity	Engineered Capacity
Line Terminations	140	160
Trunk Terminations	719	816
Matrix Load (Erlangs)	280.8	336
Processor Capability (Call Attempts/Hour)	5,616	6,720
RSH-3, Squier Hall		
Line Terminations	280	280
Trunk Terminations	28	28
Matrix Load (Erlangs)	21.8	26
Processor Capability (Call Attempts/Hour)	436	523

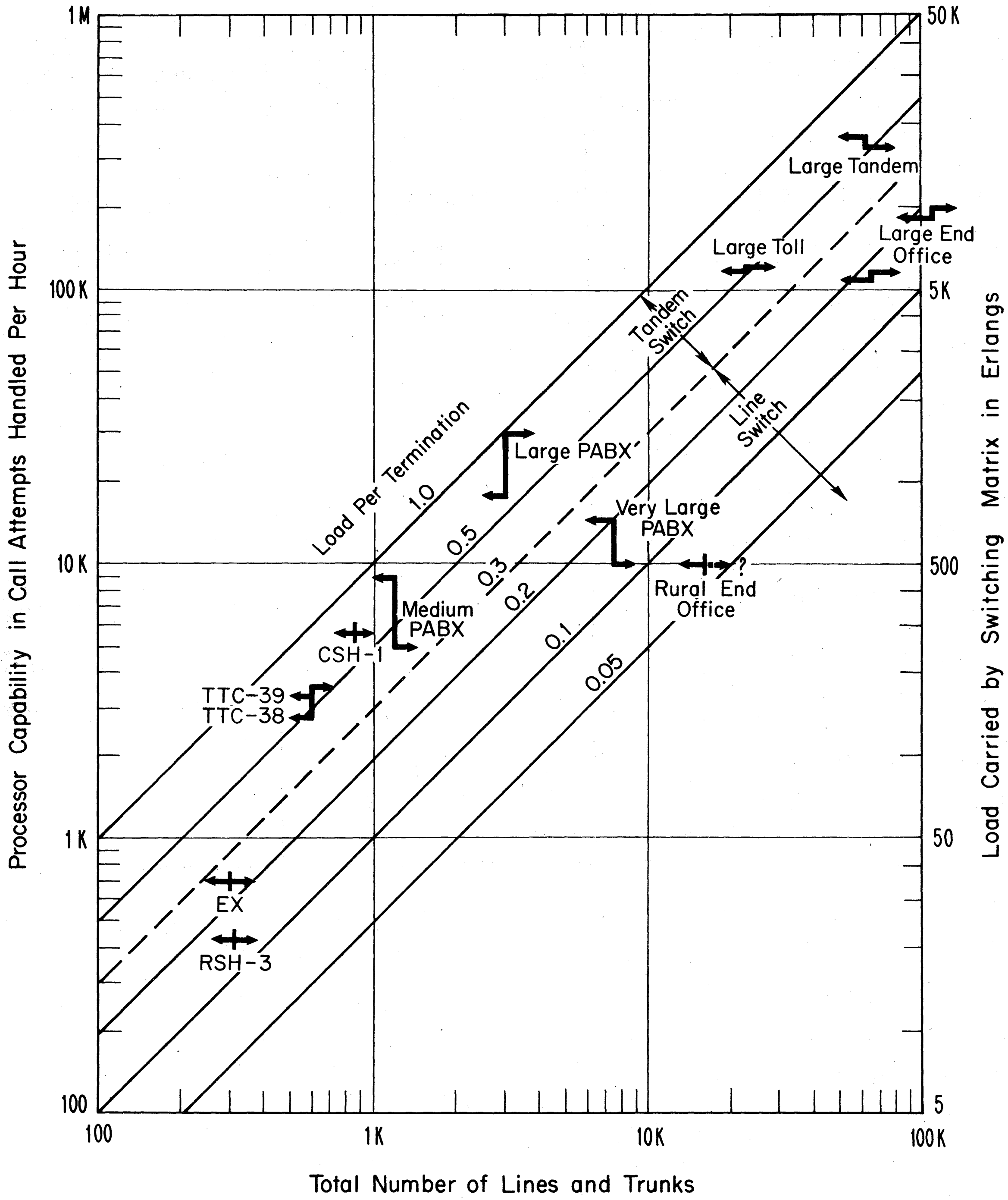


Figure 59. Relationship between switch capacity requirements assuming 3-minute average holding time.

The switch parameters for this example are also plotted on Figure 59 and designated EX. The average load per termination (lines and trunks) is found to be approximately 0.23 Erlangs.

7. SUMMARY, CONCLUSIONS AND RECOMMENDATIONS

This report has introduced, categorized, and partly defined the various switching concepts which are in use today. It has defined the basic elements of a circuit switch and described the functions the switch performs (Sec. 2). The pertinent characteristics of several existing switching systems have been tabulated (Sec. 3). With this background, the factors limiting the capacities of a digital circuit switch have been presented. The requirements on switching system's capacity are found to be a multi-dimensional quantity involving many elements of the switch. For a digital switch with stored program control, these quantities are related to traffic, terminals and control processes. In particular, in determining grade of service one must emphasize: 1) the number of terminations served; 2) the traffic load or intensity carried by the switch; 3) the number of busy hour call attempts (BHCA) handled by the processor; and 4) the control signaling subsystem (Sec. 4).

The subject of teletraffic engineering has been introduced. This is an important consideration in the planning and conceptual development of all the network subsystems. The usual grade of service parameter is probability of blocking. It has been explored for typical situations that arise in digital switching and concentration nodes. Applicable formulas were reviewed, so that concentration ratios and switch blocking probabilities can be evaluated as required. There are, unfortunately, several issues and unsolved problems in this area. Their access area impact is partly understood (Sec. 5).

Finally, the planning of switch requirements for a specific access area has been outlined. The approach, applied to Ft. Monmouth, specifies the required switching capabilities for this post. Certain simplifying assumptions are made. As applied herein, the approach generates the required capacity information for the design of a new switching system or for modifications of an existing switching system.

It is concluded that a relatively simple approach may be used for switch system planning in a given access area. The key prerequisite is that the communications profile be known, or that it can be estimated with reasonable accuracy.

Other key factors which limit the processor's capacity in a stored program control switching system are:

<u>Number and Types of Terminations</u>. This includes the data and voice termination mix, class and service feature mix, and the type of call mix, such as busy, misdialed and completed.

<u>Processor Cycles per Call</u>. This determines the average amount of time used per call. The fewer cycles required, the greater the processor capacity.

<u>Program Organization</u>. The software code and programming language used for the processor affects the average number of cycles per call and thus the capacity.

<u>Performance</u>. Network blocking affects processor capacity due to two reasons. Some time is wasted in futile attempts. Another fraction of time is spent searching for alternate routes (if such option is offered). Storage queues, which smooth traffic surges, can reduce processor capacity requirements, but only if delays can be tolerated. Even with no queueing, some capacity margin is needed for peak traffic periods.

<u>Architecture</u>. Processor capacity depends on the method used to allocate calls. If more than one processor can share the load during peak operating periods, the individual processor capacity requirements are reduced.

<u>Administrative Functions</u>. The time available to process calls depends on the overhead time required to perform administrative tasks.

<u>Growth</u>. Some processor capacity reserve is often required for future growth.

<u>Signaling System</u>. The processor may be limited by the control signaling methodology. On the line side, this limitation is usually due to the terminal signaling technique; e.g., the address number dialer. On the trunk side, common-channel signaling techniques appear to be a less severe limitation.

We have shown how these interactive factors affect the total system capacity. In some instances, only qualitative results have been obtained.

In other important instances, quantitative results have been derived. In the significant area of traffic engineering, delay systems were not covered due to time and space limitations. However, blocking criteria pertinent to digital concentration and switching have been introduced and some important delay parameter issues described. Since delay effects are important by-products of queueing systems, much work remains to be done in this area, before detailed system specifications can be developed.

When the concepts and definitions developed in the earlier sections of the report are applied to a specific access area, many of the results and objectives are clarified. The Ft. Monmouth access area is used to demonstrate how switching needs can be determined for that area. The course indicated in this report shows how further analysis and planning can be undertaken. Of course, more refinements are needed before a complete Access Area Digital Switching System can be specified, procured and implemented.

Some additional studies required for the AADSS development are outlined below.

1. Establish the non-tactical communications requirements for special features and switch functions, reliability, availability, security, and other measures of effectiveness. Apply results to the planning approach.
2. Extend switch analysis to provide for growth projections. Include digital data traffic. Integrate switch specifications and derive traffic models for various mixes of data and voice traffic.
3. Include larger regions (e.g., entire New Jersey area) and multi-service traffic statistics, to extend the analysis beyond the local access area. Include blocking probability engineering models for a number of expected communities of interest.
4. Develop traffic analysis specifically tailored for military access area. Establish tolerable delays for various classes of service and evaluate switch capacity objectives in terms of such acceptable delays. A good example is the access time requirement and how it is affected by a common signaling channel. Some other unexplored issues of interest to AADSS are indicated in Section 5.4. The queueing model described in Section 5.3 also remains to be applied to the AADSS.

5. Include cost as a parameter in development of effective access area switching hubs.

8. REFERENCES

AT&T (1961), Switching Systems (American Telephone and Telegraph Company, New York, NY).

Bear, D. (1976), Principles of Telecommunication - Traffic Engineering (P. Peregrinus Ltd., on behalf of IEE, Telecomm. Series 2, Stevenage, England).

Brand, J.E., and J.C. Warner (1977), Processor call carrying capacity estimates for stored program control switching systems, Proc. IEEE, 65, pp. 1342-1349.

CCITT (1977a), Line transmission, Vol. III, CCITT Sixth Plenary Assembly (Orange Books), ITU, Geneva, Switzerland.

CCITT (1977b), Telephone operation quality of service and tariffs (Orange Books), vol. II.2 and Supplements, ITU, Geneva, Switzerland.

Fischer, M.J., and T.C. Harris (1976), A model for evaluating the performance of an integrated circuit - and packet-switched multiplex structure, IEEE Trans. on Comm., COM-24, February, pp. 195-202.

Forgie, J.W., and A.G. Nemeth, (1977), An efficient packetized voice/data network using statistical flow control, Conference Record of International Communication Conf. (ICC '77), Chicago, IL, pp. 38.2-44/ 38.2-48.

Fry, T.C. (1928), Probability and Its Engineering Uses (D. VanNostrand Company, New York, NY).

Gitman, L., H. Frank, B. Occhiogrosso, and W. Hsieh (1977), Issues in integrated switching network design, Conference Record of International Conference on Communications, (ICC '77), Chicago, IL, pp. 38.1-36/ 38.1-43.

Harr, J.A. (1968), The design and production of real-time software for electronic switching systems, Proceedings of the NATO Software Engineering Conference, Garmish, West Germany, pp. 25-30.

Hill, M.T., and S. Kano (1976), Programming Electronic Switching Systems, (Peter Peregrinus Ltd., Southgate House, Stevenage, Hertz. Sgl 1HQ, England).

Joel, A. (1977), What is telecommunications circuit switching, Proc. IEEE, 65, pp. 1237-1253.

Kleinrock, L. (1975), Queueing Systems, Volume I: Theory (John Wiley and Sons, New York, NY).

Kleinrock, L. (1976a), ARPANET lessons, Conference Record of International Communications Conference (ICC '76), Philadelphia, PA, pp. 20-1/20-6.

Kleinrock, L. (1976b), Queueing Systems, Volume II: Computer Applications (John Wiley and Sons, New York, NY).

Kosten, L. (1948-1949), On the validity of the Erlang and Engset loss-formulae, Het. P.T.T. Bedrift, vol. 2, pp. 42-45.

Kuczura, A. (1973), The interrupted Poisson process as an overflow process, Bell Syst. Tech. J., vol. 52, pp. 437-448.

Leaky, D. (1977), Switching in space and time, Proc. IEE, 124, pp. 17-24.

Linfield, R.F. (1979), Control signaling in a military switching environment, NTIA Report 79-13, U.S. Dept. of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences, Available from NTIS, Springfield, VA 22161, Accession No. ________________.

Linfield, R.F., and M. Nesenbergs (1978), Access area signaling and switching: concepts, issues and alternatives, NTIA Report 78-2, U.S. Dept. of Commerce, National Telecommunications and Information Administration, Institute for Telecommunication Sciences. Available from NTIS, Springfield, VA 22161, Accession No. PB 283-464/AS.

Lotze, A., and W. Wagner (1963), Tables of the modified Palm-Jacobaeus-loss formulae, Institute of Switching and Data Technics, Technical University, Stuttgart, Germany.

Morse, P.M. (1958), Queues, Inventories and Maintenance (John Wiley and Sons, New York, NY).

Nesenbergs, M. (1979), A hybrid of Erlang B and C formulas and its applications, IEEE Trans. on Comm., COM-27, January, pp. 59-68.

Nesenbergs, M., and R.F. Linfield (1976), Parametric cost alternatives for local digital distribution systems, OT Report 76-95, U.S. Dept. of Commerce, Office of Telecommunications (now National Telecommunications and Information Administration), Institute for Telecommunication Sciences. Available from NTIS, Springfield, VA 22161, Accession No. PB258-510/AS.

Nowak, J.S. (1976), No. 1A-ESS, A new high capacity switching system, International Switching Symposium, Kyoto, Japan, pp. 131-1.1/131-1.7.

Palm, C. (1937), Etude dés delais d'attente, Ericsson Technics, Vol. 5, pp. 39-56.

Pitroda, S.G. (1977), Switching systems, Telecommunications, vol. 11, no. 13, pp. 75-86.

Riordan, J. (1962), Stochastic Service Systems (John Wiley and Sons, New York, NY).

Roberts, L.G. (1977), Development of packet switching networks worldwide, International Seminar on Digital Communications, Zurich, Switzerland.

Ross, M.J., A.C. Tabbot, and J.A. Warte (1977), Design approaches and performance criteria for integrated voice/data switching, Proc. IEEE, 65, September, pp. 1283-1295.

Saaty, T.L. (1961), Elements of Queueing Theory (McGraw-Hill Book Co., New York, NY).

Schwartz, M. (1977), Computer Communications Network Design and Analysis (Prentice Hall, Inc., Englewood Cliffs, NJ).

Siemens (1970), Telephone traffic theory tables and charts, Siemens, AG, Munchen, Germany.

Telesis (1976), Special issue on the SL-1 business communication system, Bell Northern Research Ltd., Ottawa, Canada, vol. 4, no. 3.

Vaulot, E. (1927), Extension des formulas d'Erlang en cas ou les durées des conversations suivent une loi quelconque, Rev. Gen. d'Elect., vol. 22, pp. 1164-1171.

Vena, P.A., and G.J. Coviello (1975), Integration of circuit/packet switching by SENET (slotted envelope network) concept, Conference Record of the National Telecommunications Conference (NTC '75), New Orleans, LA, December, pp. 42-13/42-17.

Wagner, L.H. (1977), Access area digital switching system, Proceedings of Second Annual Seminar, Armed Forces Communications and Electronics Association, Ft. Monmouth, NJ, pp. 39-57.

Wilkinson, R.I. (1956), Theories for toll traffic engineering in the USA, Bell Syst. Tech. J., vol. 35, pp. 421-514.

APPENDIX A. DETAILS OF THE CONCENTRATOR/SWITCH SERVER TRUNK PROBLEM

A.1. Problem Definition

Assume the m-server facility of Figure 39. The number of users or sources is M>m. To expedite matters, consider only exponential distributions for both interarrival times and for service (holding) times. Let blocked calls be lost or cleared without any aftereffect. Let the average service times be unity for both distant and local calls. However, let their average arrival rates be different. Denote by a the "distant" call arrival rate per user, and by b the "local" arrival rate per user. See Figure 39 to emphasize that distant calls employ a single server and local calls employ two servers. Finally, assume that the concentrator and the switch have full availability and that the switch is intrinsically non-blocking (i.e., it has no processor or real time limitations).

With this convention, one has construed a queueing system of the M/M/m/O/M type (Kleinrock, 1975), where O refers to the total lack of storage in the postulated loss system. Such systems are known to exhibit transient and equilibrium behavior (Saaty, 1961; Riordan, 1962; Kleinrock, 1975). Here, one is mainly interested in the steady state or equilibrium states and their probabilities.

Let the integer pair (i,j) stand for the joint event (or state) that i distant calls and j local calls are in progress. Then not only $0 \le i \le m$ and $0 \le 2j \le m$, but also $0 \le i+2j \le m$ must be true. Let the equilibrium probability of the state (i,j) be p(i,j). The state probabilities are needed to determine blocking probabilities for the queueing system at hand. The usual way to find p(i,j)'s is through the flow rate conservation or statistical equilibrium difference equations for the system (Saaty, 1961; Riordan, 1962; Benes, 1965; Kleinrock, 1975). The equations are constructed by simply equating the OUT and IN flows at each (i,j) state. It helps, however, to first dispose of unneeded trivial terms. To this end, define

$$p(i,j) = 0 \quad \text{for all } i<0, \quad \text{for all } j<0, \quad \text{and for all } i+2j>m. \tag{A-1}$$

Also assume for simplicity, that m, the number of server trunks, is an even number.

For each (i,j) state not trivially covered by (A-1), the "OUT=IN" equilibrium equations are:

$$[(M-i-2j)(a+b)+i+j]p(i,j)$$
$$= (M-i-2j+1)ap(i-1,j) + (i+1)p(i+1,j)$$
$$+ (M-i-2j+2)bp(i,j-1) + (j+1)p(i,j+1),$$
$$\text{if } i+2j \le m-2;$$

$$[(M-i-2j)a+i+j]p(i,j) \tag{A-2}$$
$$= (M-i-2j+1)ap(i-1,j) + (i+1)p(i+1,j)$$
$$+ (M-i-2j+2)bp(i,j-1),$$
$$\text{if } i+2j = m-1;$$

$$(i+j)p(i,j) = (M-i-2j+1)ap(i-1,j)$$
$$+ (M-i-2j+2)bp(i,j-1),$$
$$\text{if } i+2j = m.$$

There are a total of

$$\sum_{n=0}^{m/2}(2n+1) = (\tfrac{m}{2}+1)^2 \tag{A-3}$$

linear equations in (A-2), which is the same as the number of unknown p(i,j). As is well known, this set of equations is linearly dependent. One of the equations may be replaced by the standard normalization,

$$\sum_{\text{all } i,j} p(i,j) = 1. \tag{A-4}$$

To solve (A-2) and (A-4) for the $(\frac{m}{2}+1)^2$ unknowns may be a considerable problem. Consider, for instance, a T1 line spanning given concentrator-switch links. The T1 can accommodate 24 PCM lines. Thus, m=24 may be of practical interest. But, of course, $(\frac{24}{2}+1)^2=169$ implies a matrix inversion problem where the dimensionality is 169. For T2=4xT1, the dimensionality of the matrix is 2401. Furthermore, a single inversion would provide a solution only for a selected set of values, a, b, M, and m. Thus, an exact solution to this problem is not at all easy.

To proceed, one may attempt to approximate the above exact problem by a similar, more tractable, problem. In particular, one seeks here an approximating model that is indistinguishable from the original exact problem for the three limiting cases: $a \to 0$, $b \to 0$, and $M \to \infty$. Such an approximating approach is possible, as follows.

A.2. Approximating Model

Consider the service facility shown in Figure A-1. One proposes to use this model as an approximation to the true physical facility of Figure 39.

Note, first, that for the same (i,j) state the model of Figure A-1 has either the same number or more idle sources. Its effective load, as well as its blocking probability, must then be the larger of the two. The new model therefore provides a conservative design approach to the original model of Figure 39.

Second, and this shall become clear subsequently, the asymptotic (i.e., $a \to 0$, $b \to 0$, $M \to \infty$) properties for the state and blocking probabilities are by and large indistinguishable for the two models.

Third, the new model is tractable. That is, closed form expressions are possible for all the probabilities of interest.

Instead of p(i,j), let the "approximating" state probabilities be $\tilde{p}(i,j)$. In accordance with Figure A-1, the $\tilde{p}(i,j)$ are not solutions of (A-2). Rather, they must satisfy an entirely new set of "approximating" equilibrium equations:

$$[(M-i-j)(a+b)+i+j]\tilde{p}(i,j)$$
$$= (M-i-j+1)a\tilde{p}(i-1,j) + (i+1)\tilde{p}(i+1,j)$$
$$+ (M-i-j+1)b\tilde{p}(i,j-1)+(j+1)\tilde{p}(i,j+1),$$
$$\text{if } i+2j \le m-2;$$

$$[(M-i-j)a+i+j]\tilde{p}(i,j) \qquad \text{(A-5)}$$
$$= (M-i-j+1)a\tilde{p}(i-1,j) + (i+1)\tilde{p}(i+1,j)$$
$$+ (M-i-j+1)b\tilde{p}(i,j-1),$$
$$\text{if } i+2j = m-1;$$

[continued]

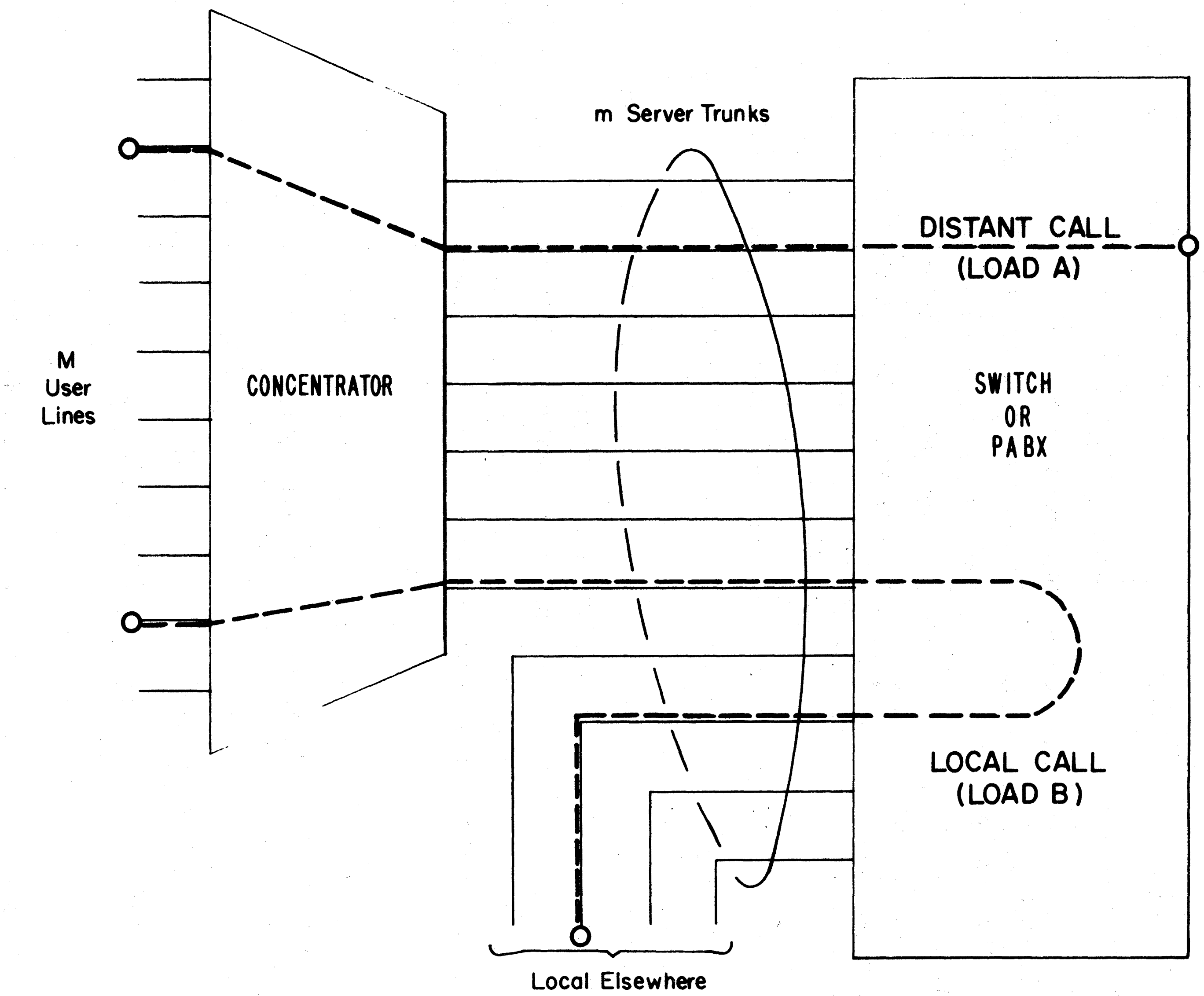


Figure A-1. An approximate model for m server trunks between a concentrator and a switch.

$$(i+j)\tilde{p}(i,j) = (M-i-j+1)a\tilde{p}(i-1,j) + (M-i-j+1)b\tilde{p}(i,j-1), \quad \text{if } i+2j = m.$$

There are also $(\frac{m}{2}+1)^2$ equations and $(\frac{m}{2}+1)^2$ unknowns here. The equations are linearly dependent and require the use of (A-4). The key difference is that on both sides of the equalities, the factor (M-i-2j) in (A-2) has been replaced by (M-i-j) in (A-5).

The asymptotic affinity of the two problems is closely related to the Engset distribution (Riordan, 1962; Bear, 1976), especially as it applies to finite number of sources M. To simplify, assume the following notation for Engset distribution:

$$\dot{E}_n(m;M,a) = \binom{M}{n}a^n\left[\sum_{k=0}^{m}\binom{M}{k}a^k\right]^{-1}, \qquad \text{(A-6)}$$

where n=0,1,2,...,m.

Then, when the traffic tends to be all local and $a\to 0$, the exact solution gives

$$p(i,j) \to E_j(\tfrac{m}{2};\tfrac{M}{2},2b) \qquad \text{if } i = 0,$$
$$\to 0 \qquad \text{if } i \neq 0. \qquad \text{(A-7)}$$

As $a\to 0$, the approximate solution tends to

$$\tilde{p}(i,j) \to E_j(\tfrac{m}{2};M,b) \qquad \text{if } i = 0,$$
$$\to 0 \qquad \text{if } i \neq 0. \qquad \text{(A-8)}$$

The total offered load in both cases is Mb. For M reasonably large, the load is the dominant factor that determines the value of $E_j(.;.,.)$. Thus, $p(i,j)$ and $\tilde{p}(i,j)$ are approximately equal for $a\to 0$.

When the offered traffic is nearly all distant, $b\to 0$. Then the exact and approximate solutions are indistinguishable in the limit:

$$p(i,j),\ \tilde{p}(i,j) \to E_i(m;M,a) \qquad \text{if } j = 0,$$
$$\to 0 \qquad \text{if } j \neq 0. \qquad \text{(A-9)}$$

Suppose next that the number of sources is extremely large, but their offered load is finite. In the limit $M\to\infty$, let $\lim(Ma)=A_0$ and $\lim(Mb)=B_0$.

Then as $M \to B$, the two solutions tend to the same, not Engset, distribution (Bear, 1976):

$$p(i,j),\ \tilde{p}(i,j) \to \frac{\dfrac{A_0{}^i B_0{}^j}{i!\ j!}}{\sum\limits_{\text{all } s,t} \dfrac{A_0{}^s B_0{}^t}{s!\ t!}}, \tag{A-10}$$

where $i+2j \le m$ comprises the summation domain for (s,t) as well.

Because of the asymptotic properties (A-7) to (A-10), one feels justified to call $\tilde{p}(i,j)$ an approximation of the true $p(i,j)$.

A.3. The Approximate Solution

The formal solution of (A-5) can be deduced by inspection as:

$$\tilde{p}(i,j) = \frac{\binom{M}{i+j}\binom{i+j}{i} a^i b^j}{\sum\limits_{\text{all } s,t} \binom{M}{s+t}\binom{s+t}{s} a^s b^t}, \tag{A-11}$$

where the domains of (i,j) and (s,t) are again the same.

To prove that $\tilde{p}(i,j)$ of (A-11) is indeed a solution to the approximate problem, it suffices to substitute (A-11) into (A-5). We skip the lengthy details here. The main point is that $\tilde{p}(i,j)$, as given in (A-11), satisfies all three parts of (A-5).

From the approximating state probabilities $\tilde{p}(i,j)$, one proceeds to derive the blocking probabilities as functions of the offered load. To do so, one must first resolve several relatively minor difficulties with terms and their interpretation (Cooper, 1972).

First, one can define the probability of blocking as either "based on time" or as "based on traffic." If based on time, the blocking event for distant calls is tantamount to an outside observer finding the system in the state $i+2j=m$. Then, P_D, the probability of distant call blocking must be

$$P_D = \sum_{i+2j=m} p(i,j), \tag{A-12}$$

and likewise, the local call probability of blocking, P_L, must be

$$P_L = \sum_{i+2j=m-1}^{m} p(i,j). \tag{A-13}$$

These two probabilities, P_D and P_L, are said to be based in time. They are visible to a completely independent outside observer who monitors the availability of idle servers at some random times.

A user who is part of the traffic sees a somewhat different statistical situation. After all, as the number of calls in progress increases, i+2j grows, there remain fewer idle sources, and the conditional frequency of call arrivals decreases. Thus, for finite M, more traffic tends to be generated when the system is less loaded. The probabilities of blocking that are based on actual traffic percentages, blocked versus carried, therefore are slightly lower than those based on percentages of time. The following heuristic argument gives an indication how the two probabilities may be related. Imagine that one individual out of the M users is cast in the role of an "outside" observer. That individual sees, in effect, how the remaining M-1 users congest the facility. Thus, approximately,

$$\begin{aligned} P_D(M|\text{traffic}) &\cong P_D(M-1|\text{time}), \\ P_L(M|\text{traffic}) &\cong P_L(M-1|\text{time}), \end{aligned} \tag{A-14}$$

where time-tagged probabilities are the same as given in (A-12) and (A-13). For practical approximation purposes, the transformation between time and traffic based blocking probabilities is thus straightforward.

The second minor difficulty concerns the effective offered load. For well-designed systems, it is only incrementally larger than the actual carried load and poses no problem. However, the finite source assumption causes some potential confusion.

Let A be the effective total distant load, and B the effective total local load. Then $A \le Ma$ and $B \le Mb$, because the instantaneous arrival rate can only decrease as more and more calls become active. For the classical Engset distribution the situation is better understood. Thus, when b=0 (i.e., there is no local traffic and (A-9) is valid), one can prove that the effective distant offered load is

$$A = \frac{Ma}{1+a[1-E_m(m;M-1,a)]} . \tag{A-15}$$

Here, $E_m(m;M-1,a)$ is nothing more than the blocking loss observed on traffic basis.

Unfortunately, for arbitrary mixes of distant and local traffic, it is not entirely clear how (A-15) generalizes to relationships between A, B on one hand and Ma, Mb on the other. One can, by all means, use the "initial" (i.e., i=j=0) offered loads Ma and Mb whenever expedient. At other times, when effective offered loads are desired and the blocking probabilities P_D and P_L do not exceed 1%, one can approximate the effective loads as

$$A \cong Ma/(1+a),$$
$$B \cong Mb/(1+b). \qquad \text{(A-16)}$$

The transformation between initial loads, Ma and Mb, and effective loads, A and B, (A-16), is therefore straightforward.

A.4. Numerical Results

This section presents numerical calculations done for the approximate model of Figure A-1. The section evaluates blocking probabilities for the approximate solution, denoted as $\tilde{p}(i,j)$ in (A-11). These computations were done largely on a hand calculator and are rather limited in scope. Neither the curves nor the numbers should be taken as direct substitutes for the exact solution. Rather, the approximation should be interpreted as an indicator of common trends. The eventual goal of these calculations has been to resolve the uncertainties shown in Figure 40 of the main text and, if possible, to generate traffic engineering curves of the type presented in Figure 41.

The initial computed curves, see Figures A-2 and A-3, show approximate probabilities of blocking in the 10^{-1} to 10^{-3} range. This emphasizes the broadly accepted practice of seeking grade of service objectives in the neighborhood of one percent. The abscissa in both figures is the initial offered load. Both figures assume a relatively modest number of users, M=20, and a correspondingly small number of servers, m=10. Furthermore, the probabilities depicted (such as P_D in Figure A-2) are based on fraction of time blocked. To translate the results to fractions of traffic blocked, one uses (A-14). Likewise, either (A-15) or (A-16) should be used to deduce the effective offered load.

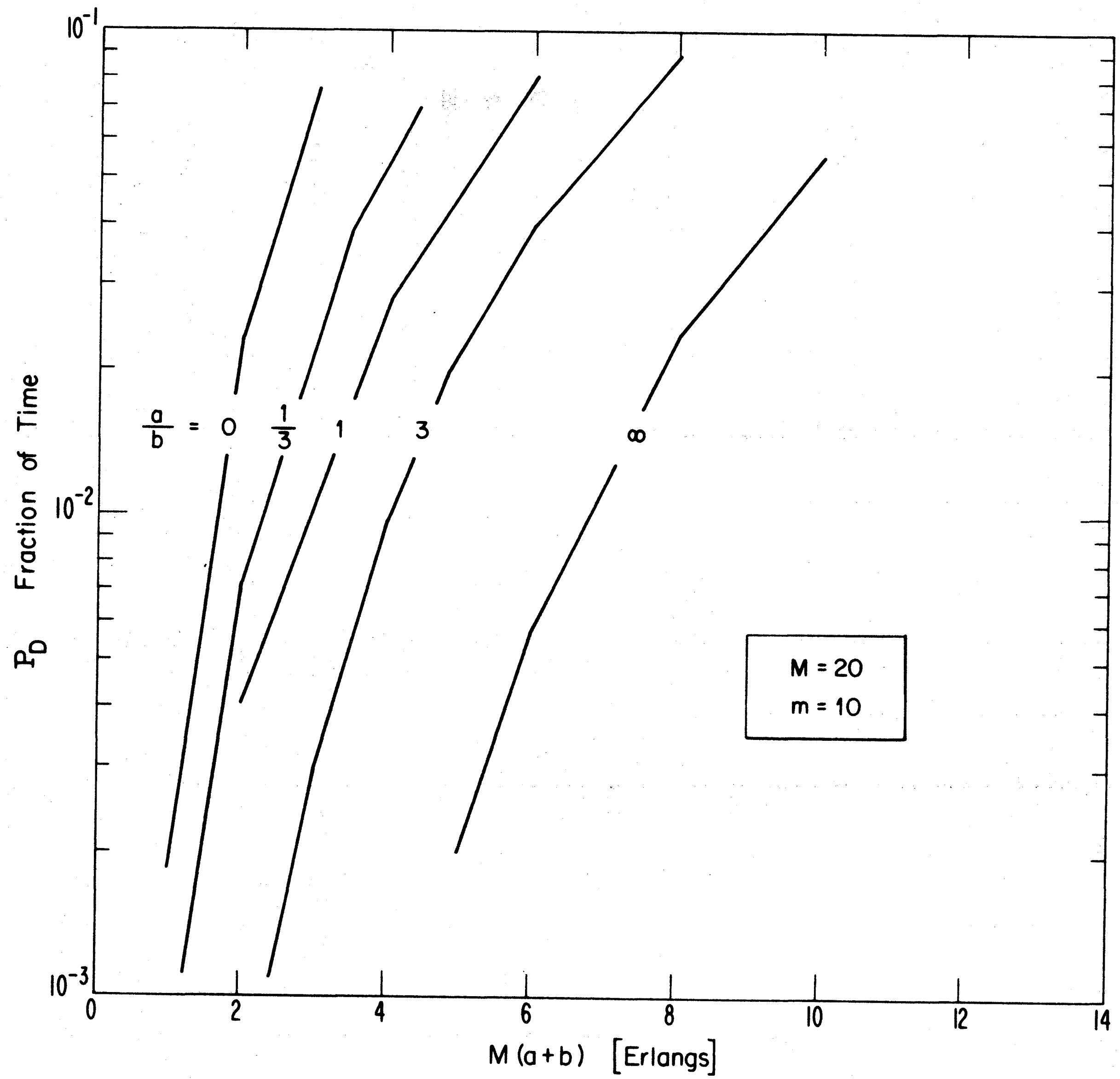


Figure A-2. Approximate probability of blocking for distant calls P_D (based on fraction of time) versus the initial load M(a+b) parametric in distant/local load ratio a/b, for M=20, m=10.

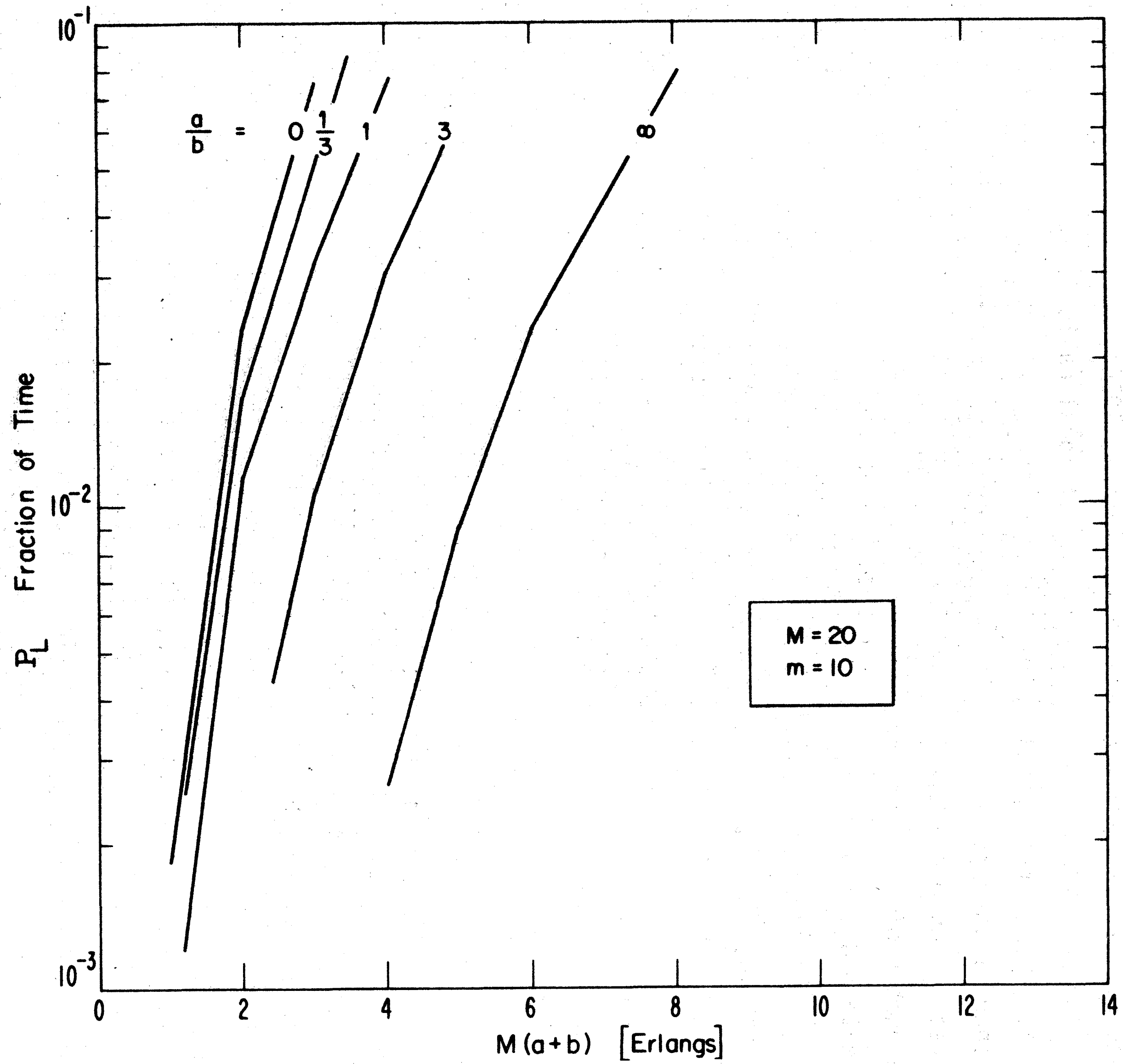


Figure A-3. Approximate probability of blocking for local calls P_L (based on fraction of time) versus the initial load M(a+b) parametric in distant/local load ratio a/b, for M=20, m=10.

For given values M, m, a, and b, the actual computation for P_D is carried out using equations (A-11) and (A-12). For P_L one uses (A-11) and (A-13).

The results of these initial computations show that both blocking probabilities, P_D and P_L, may depart appreciably from classic Erlang and Engset grade of service curves (AT&T, 1961; Siemens, 1970).

For nearly all $a/b \neq 0$, the blocking of local calls is seen to be considerably more frequent than that of distant calls. This aspect is further illustrated in Figure A-4, which directly contrasts P_D versus P_L curves. Figure A-4 differs from the previous in several other respects. The number of users (sources) and the number of server trunks have been more than doubled to M=48 and m=24.

Intuitively, one might expect P_L to be roughly twice P_D. But things are really not that simple. The ratio P_L/P_D is plotted in Figure A-5. One sees that P_L/P_D depends on the distant to local load ratio, a/b, as well as on the number of servers m. The points of Figure A-5 are derived from the same computation that led to previous Figures A-2 to A-4. The two curves drawn in Figures A-5 are simple graphical fits. As functions of m, a, and b, they are of the form

$$\frac{P_L}{P_D} \cong 1 + \frac{85}{m} \left[\frac{a}{m(a+b)}\right]^{5/m} . \qquad \text{(A-17)}$$

Note, that for the practical case a/b=1/3 (or a/(a+b)=1/4), the ratio P_L/P_D is nearly 2.4 for both m=10 and m=24. The least value of the ratio is unity. It occurs at a/b=0.

Since the computation of every single P_D and P_L point value is time consuming, the numerical effort may be reduced in half through the use of (A-17) or Figure A-5.

For a constant offered load, an increase in the number of servers, m, reduces the blocking probabilities. But, for the same constant load an increase in the number of users, M, does the opposite. It increases the blocking probabilities. However, the effect of M on P_D and P_L is less dramatic. This phenomenon is illustrated in Figure A-6 for effective offered load ratio A/B=1/3. The curves, which are parametric in M and m, are not

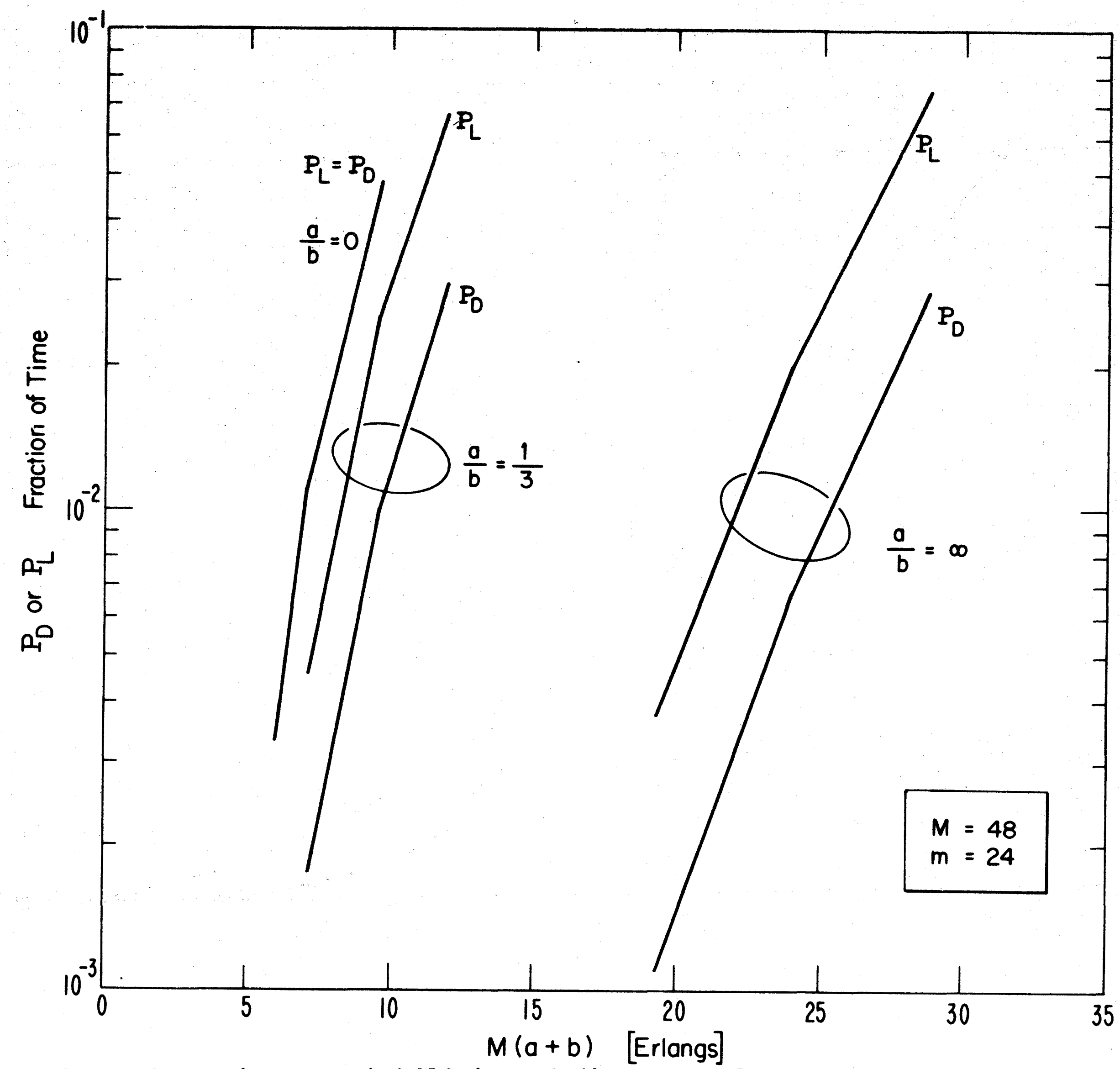


Figure A-4. Approximate probabilities of distant call blocking P_D and local call blocking P_L (both based on time) versus the initial load M(a+b), parametric in a/b, for M=48, m=24.

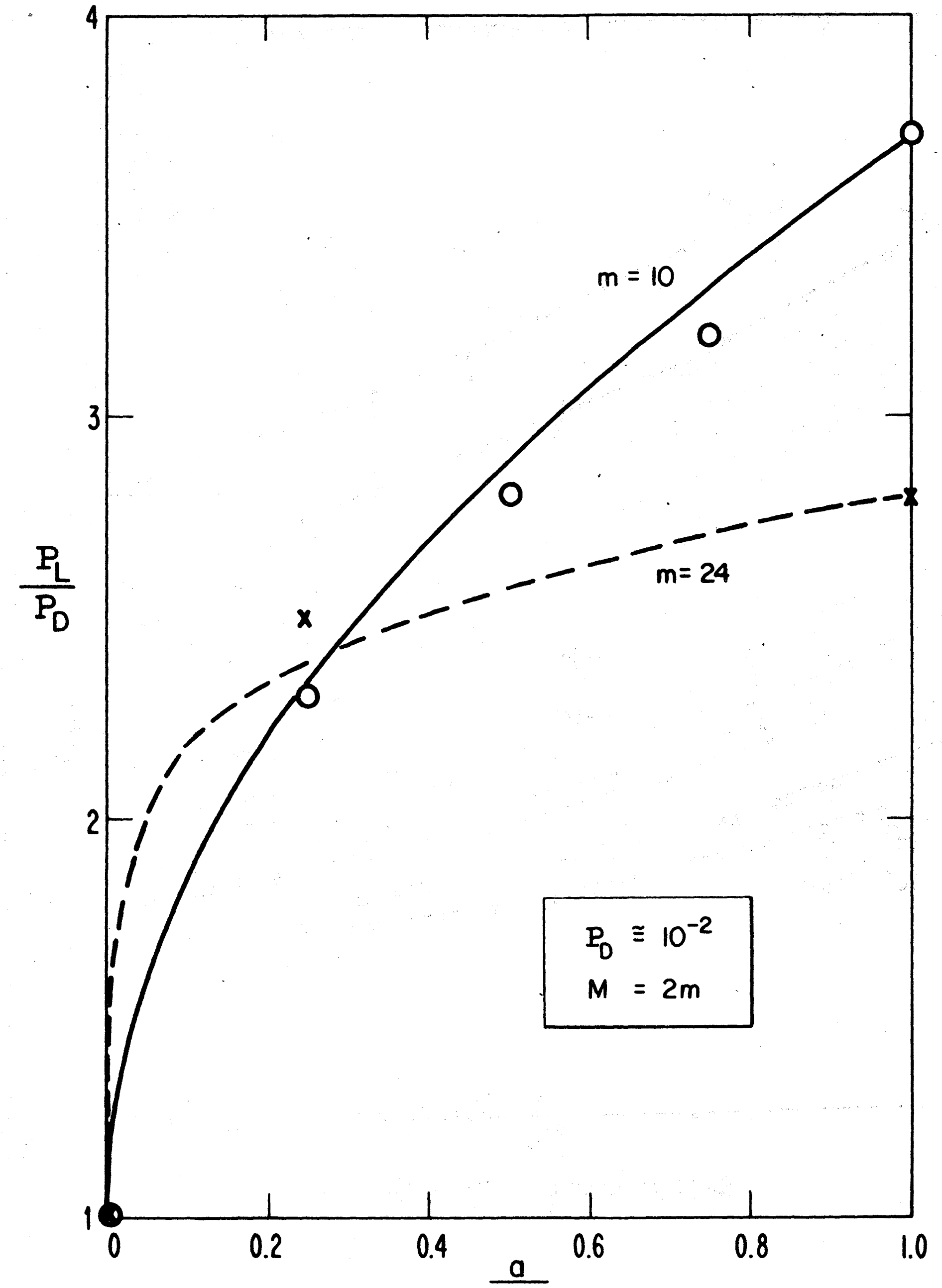


Figure A-5. The approximate ratio of local and distant call blocking probabilities, P_L/P_D, for $P_D \cong 10^{-2}$ and 2:1 concentration of lines.

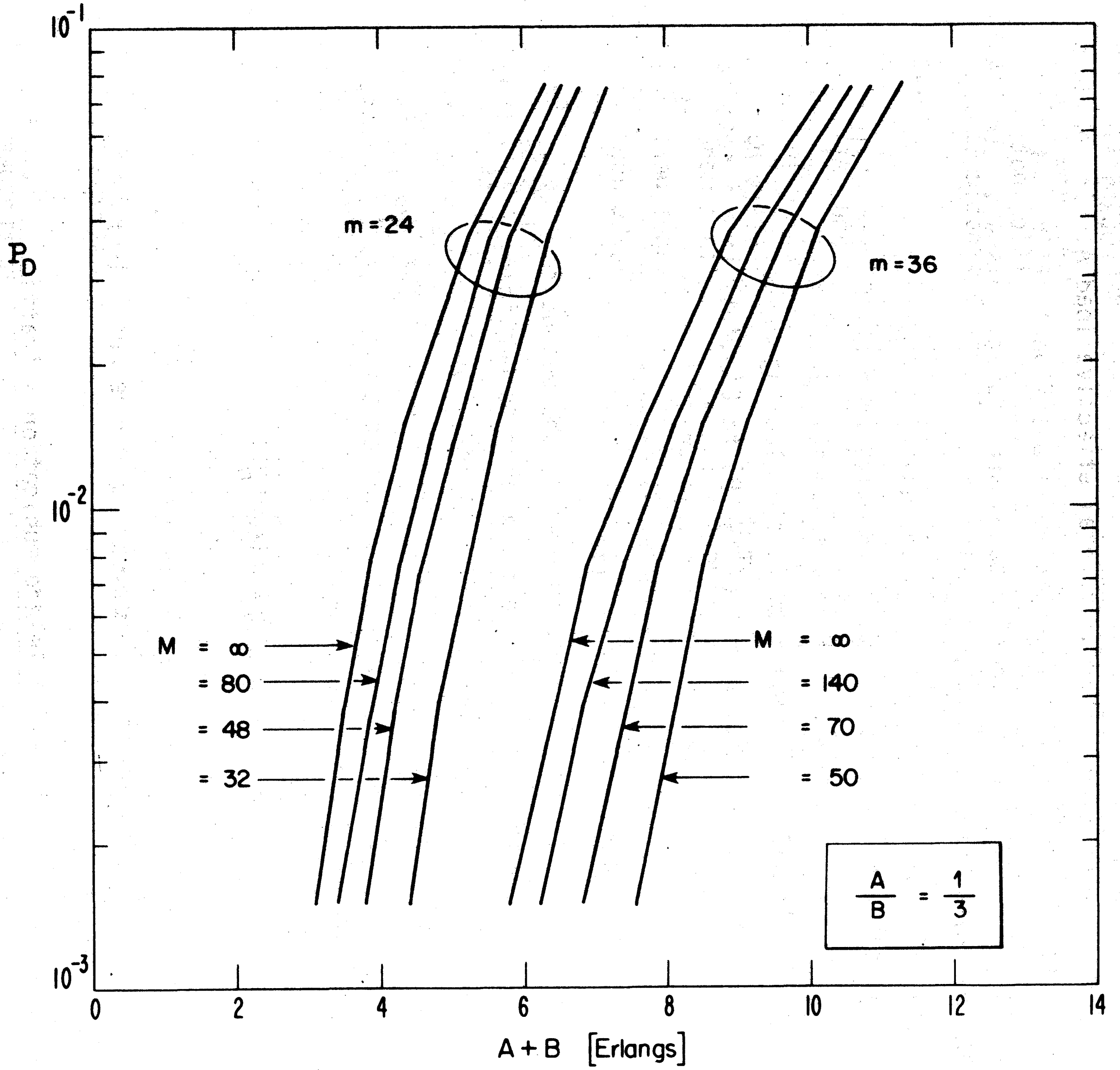


Figure A-6. The approximate distant call blocking probablity P_D versus effective total offered load, A+B, parametric in M and m, and subject to A/B=1/3.

computed on the basis of previous approximations, but instead are the end result of an even rougher estimate

$$P_D \cong \frac{A}{A+B} E(m;M,A) + \frac{B}{A+B} E(\frac{m}{2};\frac{M}{2},2B), \qquad \text{(A-18)}$$

where E(m;M,A) stands for the tabulated Engset blocking probability (Siemens, 1970) that is specified in terms of total effective load A. The grade of service itself (i.e., P_D) is interpreted as based on traffic, not time. Since for M>>1 the time and traffic bases are indistinguishable (see (A-14)), (A-18) may often be used either way.

Several other comments apply to the use of (A-18):

(i) $A/B \cong a/b$ may depart from an identity [see (A-15) and (A-16)]. However, it appears to be a reasonable approximation in practice.

(ii) As A/(A+B) goes from 0 to 1, the blocking probability P_D traverses its range from $E(\frac{m}{2};\frac{M}{2},2B)$ to E(m;M,A) in a monotonic fashion. To the extent that this transition is not perfectly uniform or linear, (A-18) must be interpreted as an approximation. Compare Figures A-2, A-3, and A-4.

(iii) Of the two Engset terms in (A-18), the second term is usually the dominant one. If the first term is negligible, then the curves akin to those given in Figure 41 of main text can be simply generated. For the assumed $P_D(=10^{-2})$ and A/B (=1/3) values, one merely performs an Engset table lookup (Siemens, 1970) of M and m that satisfy

$$E(\frac{m}{2};\frac{M}{2},2B) \cong (1+\frac{A}{B})P_D. \qquad \text{(A-19)}$$

A.5. References

AT&T (1961), Switching Systems (American Telephone and Telegraph Co., New York, NY).

Bear, D. (1976), Principles of Telecommunication - Traffic Engineering (P. Peregrinus Ltd., on behalf of IEE, Telecomm. Series 2, Stevenage, England).

Benes, V.E. (1965), Mathematical Theory of Connecting Networks and Telephone Traffic (Academic Press, New York, NY).

Cooper, R.B. (1972), Introduction to Queueing Theory (MacMillan Co., New York, NY).

Kleinrock, L. (1975), Queueing Systems, Volume I: Theory (John Wiley and Sons, New York, NY).

Riordan, J. (1962), Stochastic Service Systems (John Wiley and Sons, New York, NY).

Saaty, T.L. (1961), Elements of Queueing Theory (McGraw-Hill Book Co., New York, NY).

Siemens (1970), Telephone Traffic Theory Tables and Charts (Siemens AG, Munchen, Germany).

APPENDIX B. DETAILS OF THE COMMON FACILITY PROBLEM FOR TWO TYPES OF USERS

B.1. Problem and Its Solution

Assume the m server facility of Figures 43 and 44. There are two classes of users indicated. The first class, perhaps too loosely referred to as "lines," contains L users. Each of the lines carries an average load of $a=\lambda_1/\mu_1$ Erlangs, where λ_1 is the average arrival rate and μ_1 is the average service rate. Under the exponential service time assumption, the latter implies the average service or holding time to be $1/\mu_1$. The second class, loosely called "trunks," consists of T users, each burdened with $b=\lambda_2/\mu_2$ Erlangs per trunk. Here, of course, λ_2 and μ_2 have the same arrival and service rate interpretation as for lines. Note: These loads a and b should not be confused with the other loads, that in the main text or in Appendix A may be denoted by the same symbols. Either lines or trunks may be multiplexed or concentrated on a same bus. In that case, one observes only the total arrival rates of the two classes, such as perhaps Tb for the data packet rate [see part (C) of Figure 42, main text]. As in Appendix A, assume again exponential distributions for both interarrival and service (holding) times.

Let blocked calls undergo the random θ versus $1-\theta$ coin toss implied in Figure 44. Thus, when $\theta=1$, one has the blocked calls queued or delayed situation. When $0<\theta<1$ holds, one has a version of the generalized loss-delay hybrid (Nesenbergs, 1979).

Let the average holding times be distinct, $\mu_1 \neq \mu_2$, for the time being, for the two user classes. Let their average arrival rates also differ. Assume again full availability to all servers by all users.

Let the pair of integers (i,j) represent the event that exactly i from the lines and j from trunks are either receiving service or are waiting in queue. Then one must have $0 \leq i \leq L$ and $0 \leq j \leq T$.

Let the steady state or equilibrium probability of event (i,j) be p(i,j). As in Appendix A, these state probabilities may be determined from the flow rate conservation equations. For this problem, the "OUT=IN" equilibrium flows at each state (i,j) must be given by:

$$[(L-i)\lambda_1 + (T-j)\lambda_2 + i\mu_1 + j\mu_2]p(i,j)$$
$$= (L-i+1)\lambda_1 p(i-1,j) + (i+1)\mu_1 p(i+1,j)$$
$$+ (T-j+1)\lambda_2 p(i,j-1) + (j+1)\mu_2 p(i,j+1),$$
$$\text{if } 0 \le i+j < m;$$

$$[(L-i)\theta\lambda_1 + (T-j)\theta\lambda_2 + i\mu_1 + j\mu_2]p(i,j)$$
$$= (L-i+1)\lambda_1 p(i-1,j) + (i+1)\mu_1 p(i+1,j) \qquad \text{(B-1)}$$
$$+ (T-j+1)\lambda_2 p(i,j-1) + (j+1)\mu_2 p(i,j+1),$$
$$\text{if } i+j = m;$$

$$[(L-i)\theta\lambda_1 + (T-j)\theta\lambda_2 + i\mu_1 + j\mu_2]p(i,j)$$
$$= (L-i+1)\theta\lambda_1 p(i-1) + (i+1)\mu_1 p(i+1,j)$$
$$+ (T-j+1)\theta\lambda_2 p(i,j-1) + (j+1)\mu_2 p(i,j+1),$$
$$\text{if } i+j > m.$$

In this, it is always assumed that $p(i,j)=0$ whenever $\min(i,j)<0$.

The number of unknowns, $p(i,j)$, and the number of equations is $(L+1)(T+1)$ in (B-1). Fortunately, the exact solution can be obtained by simply postulating that the solution has a separation or product property (Cooper, 1972). We illustrate the nature of this result by working out the details of the very first of the three equations contained in (B-1). That is, for $0 \le i+j<m$, assume

$$p(i,j) = Cp_1(i)p_2(j), \qquad \text{(B-2)}$$

where C is some constant. Then the first part of (B-1) becomes

$$\mu_1 p_2(j)\{[(L-i)a+i]p_1(i) - (L-i+1)ap_1(i-1) - (i+1)p_1(i+1)\}$$
$$+ \mu_2 p_1(i)\{[(T-j)b+j]p_2(j) - (T-j+1)bp_2(j-1) - (j+1)p_2(j+1)\} = 0. \qquad \text{(B-3)}$$

Clearly, if the two {...} expressions vanish, one has a solution. Both parentheses represent Engset type of state probabilities (Kleinrock, 1975). The complete solution therefore is given by

$$\frac{p(i,j)}{p(0,0)} = \binom{L}{i}\binom{T}{j}a^i b^j \qquad \text{if } 0 \le i+j \le m,$$

$$= \binom{L}{i}\binom{T}{j}\theta^{i+j-m}a^i b^j \qquad \text{if } m \le i+j \le T+L. \qquad \text{(B-4)}$$

The p(0,0) factor can be deleted from (B-4) by the usual normalization condition that sums all p(i,j) to unity.

Direct substitution of (B-4) into (B-1) verifies that (B-4) is in fact a valid solution.

The blocking probability P_m for m≥1 servers follows readily from (B-4). Here, P_m will be interpreted as a "fraction of time" probability. To interpret it in terms of traffic percentages, the transformation discussed in the earlier Appendix A should be used.

The exact probability of blocking expression (time basis) for the two user-type problem can be written in several ways:

$$P_m = \sum_{i+j=m}^{L+T} p(i,j)$$

$$= \sum_{k=m}^{L+T} \sum_{i=0}^{k} p(i,k-i) \qquad \text{(B-5)}$$

$$= \sum_{k=m}^{L+T} S_k / \sum_{i=0}^{L+T} S_i,$$

where the finite sum S_k terms will be defined shortly. They will turn out to be convenient to alleviate the computational effort. First, however, introduce a function that is defined over all non-negative integers k:

$$\theta_m(k) = 1 \qquad \text{if } 0 \le k \le m,$$
$$= \theta^{k-m} \qquad \text{if } m \le k \le \infty. \qquad \text{(B-6)}$$

Then, in terms of this $\theta_m(k)$, S_k is defined as

$$S_k = \theta_m(k) \sum_{i=0}^{k} \binom{L}{i}\binom{T}{k-i} a^i b^{k-i}$$

$$= \theta_m(k) b^k \binom{T}{k} F(-L,-k;-T-k+1;\tfrac{a}{b}), \qquad \text{(B-7)}$$

where F(i,j;k;x) or ${}_2F_1(i,j;k;x)$ is the Gauss hyptergeometric series (Abramowitz and Stegun, 1964):

$$F(i,j;k;x) = \sum_{n=0}^{\infty} \frac{(i)_n (j)_n}{(k)_n} \frac{x^n}{n!} \quad . \tag{B-8}$$

For negative i or j, the Gauss series reduces to a polynomial. This makes the evaluation of equation (B-7) a finite task.

B.2. Numerical Results

Blocking probability P_m has been computed for several values of parameters m, θ, L, T, a and b, using the exact solution as given in equations (B-5) to (B-8). A sample of the computations is shown in Figure B-1. The figure plots P_m (fraction of time) versus the initial offered load, La+Tb. To transform the x-axis into the slightly lower values of the "effective offered load," one has to carry out the modifications indicated in Appendix A.

The exact points of Figure B-1 apply to rather modest parameter values; namely, to m=5, θ=0, L=50, T=10, and b=.1. The load per line, a, is varied to achieve the range of initial loads plotted. The range of P_m values is purposely kept around the 1% level because that appears to be of practical interest to access area applications. Since the evaluation of even such simple points turns out to be quite time consuming, other quicker methods of finding P_m must be of interest.

The two approaches taken here involve bounds and approximations. For an upper bound on P_m, one may take for all $0 \le \theta \le 1$ the Erlang B and C hybrid (Nesenbergs, 1979):

$$P_m \le \frac{B(m,La+Tb)C(m,La+Tb)}{\theta B(m,La+Tb)+(1-\theta)C(m,La+Tb)} \quad . \tag{B-9}$$

This bound is particularly easily evaluated with the available Erlang B tables (AT&T, 1961; Siemens, 1970), if written in an asymmetric way,

$$\frac{1}{P_m} \ge \frac{\theta}{m}(La+Tb) + \frac{1-\frac{\theta}{m}(La+Tb)}{B(m,La+Tb)} \quad . \tag{B-10}$$

Note, that for θ=0, one always has $P_m \le B(m,La+Tb)$.

This upper bound on the previously computed true values is shown in Figure B-2. The upper bound appears to be nearly 50% above the exact points for this case. It seems too unprecise for approximation purposes.

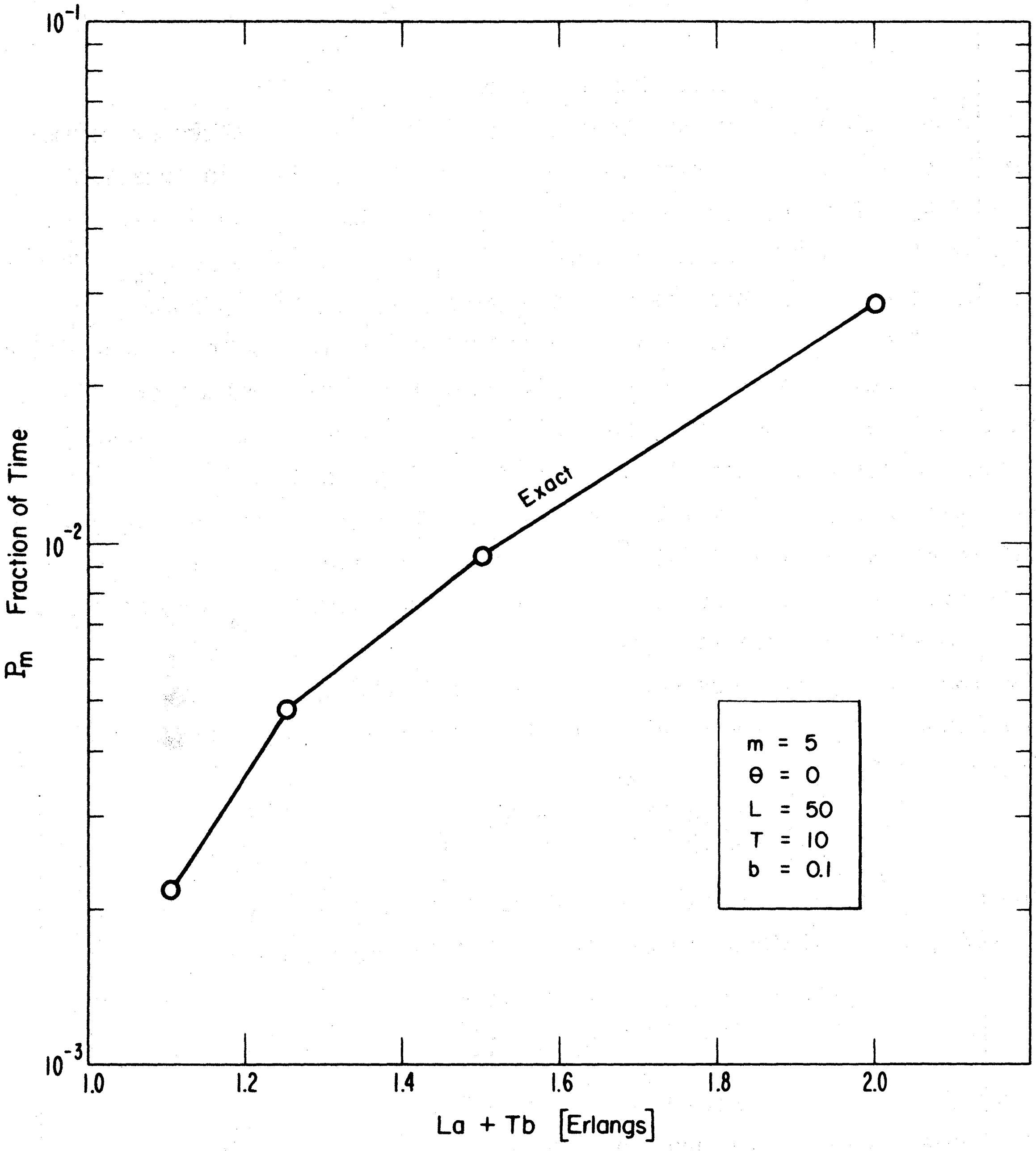


Figure B-1. Probability of blocking P_m (based on fraction of time) versus the initial load, La+Tb, for parameter values indicated.

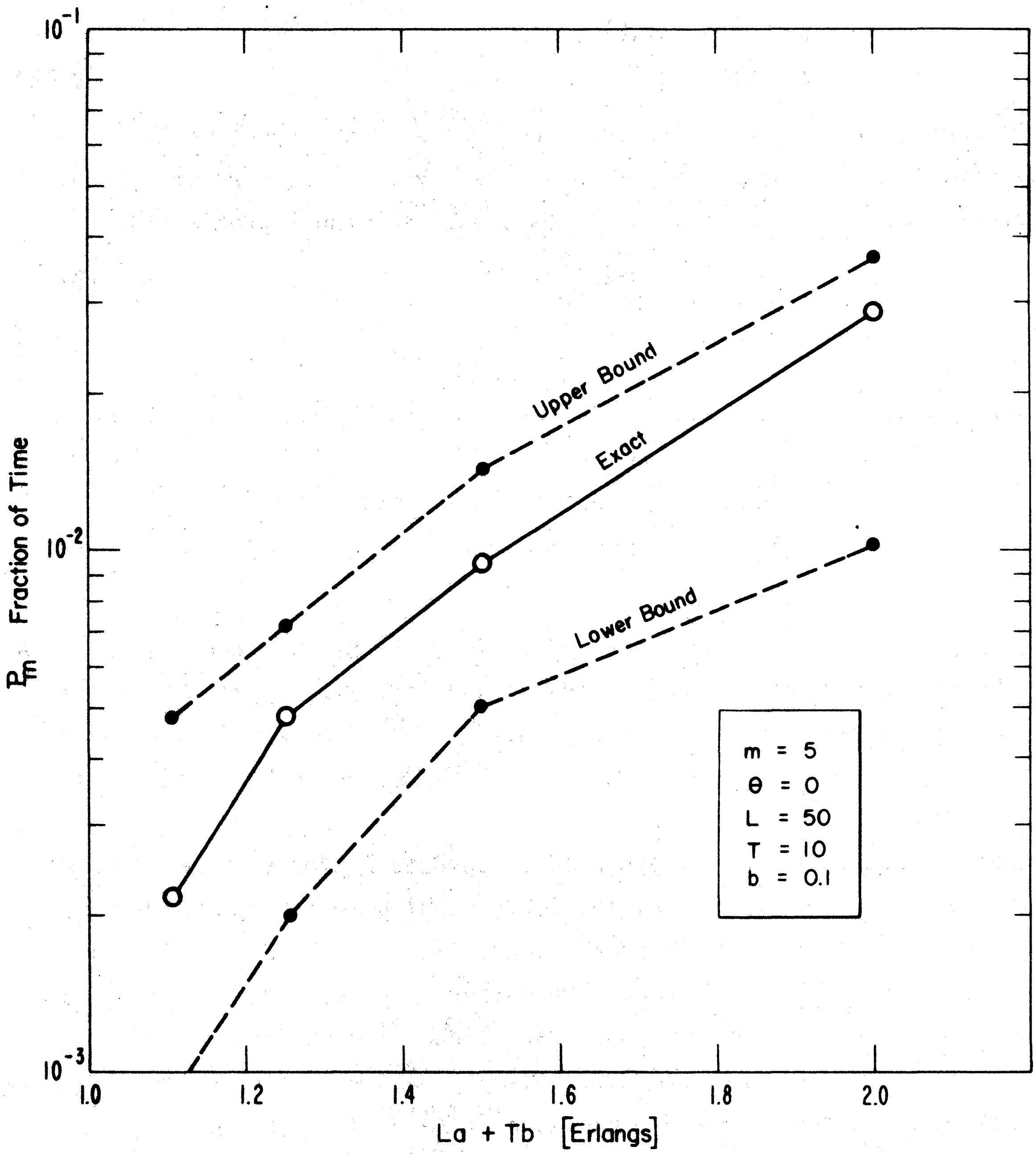


Figure B-2. Bounds on probability of blocking P_m (based on fraction of time) versus the initial load, La+Tb, for parameter values indicated.

Figure B-2 also depicts a lower bound on P_m. The lower bound is again off by some 50%. It happens to be based on the formula

$$P_m \geq e^{-(La+Tb)} \binom{L}{I} \binom{T}{m-I} a^I b^{m-J}, \qquad \text{(B-11)}$$

where in the nearest integer sense

$$\begin{aligned} I &\cong mLa/(La+Tb), \\ J &\cong mTb/(La+Tb). \end{aligned} \qquad \text{(B-12)}$$

An approximate value for P_m could be derived from averaging the two bounds (Fig. B-2) in any of several possible ways. This, however, is not the approach taken here. In what follows, one uses the approximation

$$P_m \cong A_0(1+\alpha_1\theta+\alpha_2\theta^2), \qquad \text{(B-13)}$$

where

$$\begin{aligned} A_0 &= e^{1-(La+Tb)} \binom{L}{I} \binom{T}{m-I} a^I b^{m-J}, \\ \alpha_1 &= \frac{1}{2}\left(\frac{L-J}{I+1}\, a + \frac{T-J}{J+1}\, b\right), \\ \alpha_2 &= \frac{(L-I)(T-J)}{(I+1)(J+1)}\, ab. \end{aligned} \qquad \text{(B-14)}$$

When θ is negligibly small, the approximation (B-13) reduces to e=2.718 times the previous lower bound [see (B-11) and Fig. B-2]. The fit of the approximation to the exact values is illustrated in Figure B-3. In the 1% neighborhood for P_m, one expects this approximation to be within a $\pm$10% relative error range.

Approximation (B-13) has been used to compute P_m for various service facility configurations. For initial loads ranging from 2 to 10 Erlangs, and for θ=0, 1/2, 1, this is shown in Figure B-4. Figure B-4 again plots probability of blocking, P_m, versus the number of servers m. The range of P_m is restricted to a relatively short and linear region from .5% to 2%. As seen, the effect of the hybrid loss-delay parameter θ becomes increasingly noticeable only as the load and the number of servers increase.

The loci at which P_m assumes a selected constant value, such as 1%, can be readily interpolated from Figure B-4. When this is done for

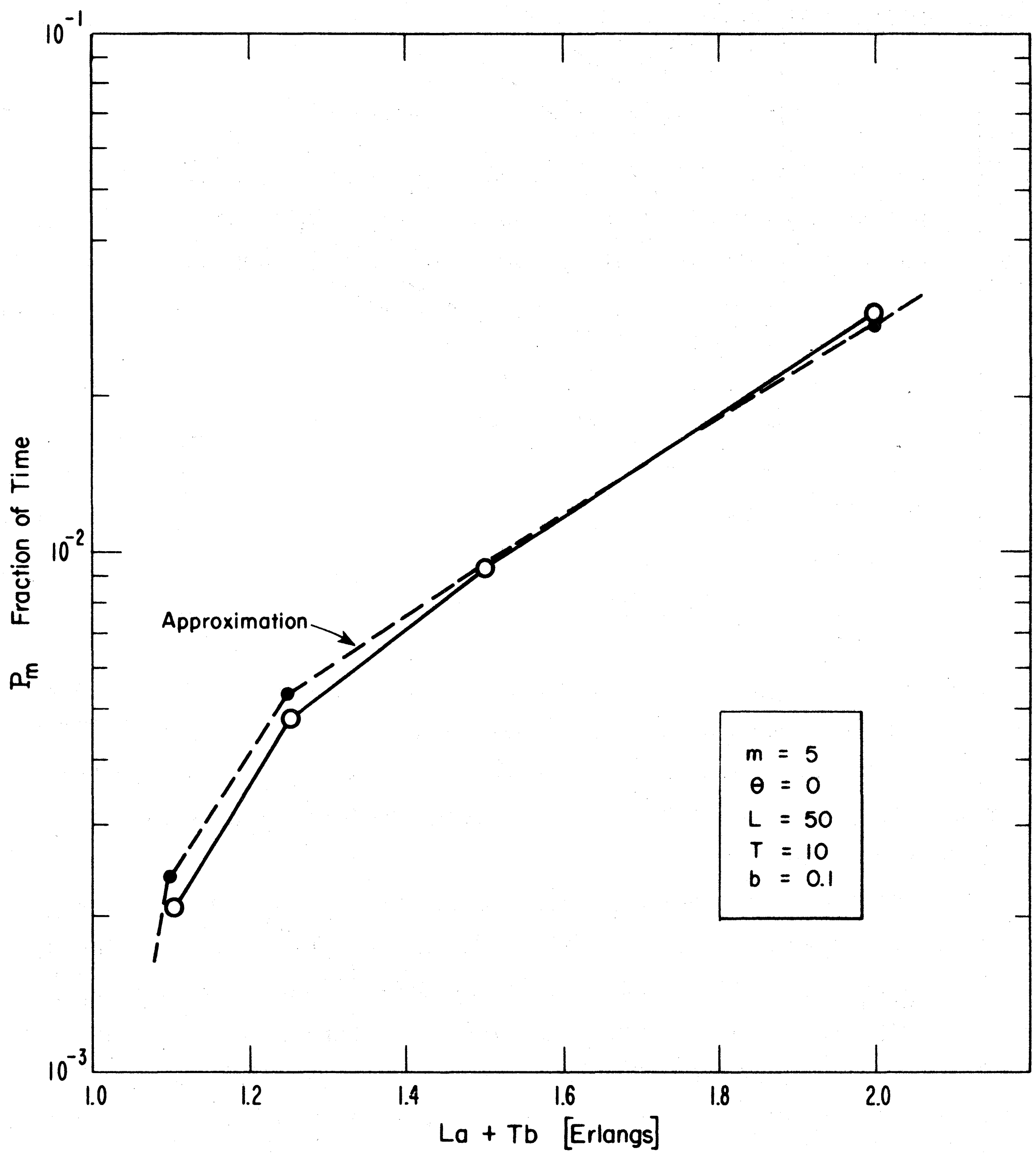


Figure B-3. Approximation of blocking probability P_m (based on fraction of time) versus the initial load, La+Tb, for parameter values indicated.

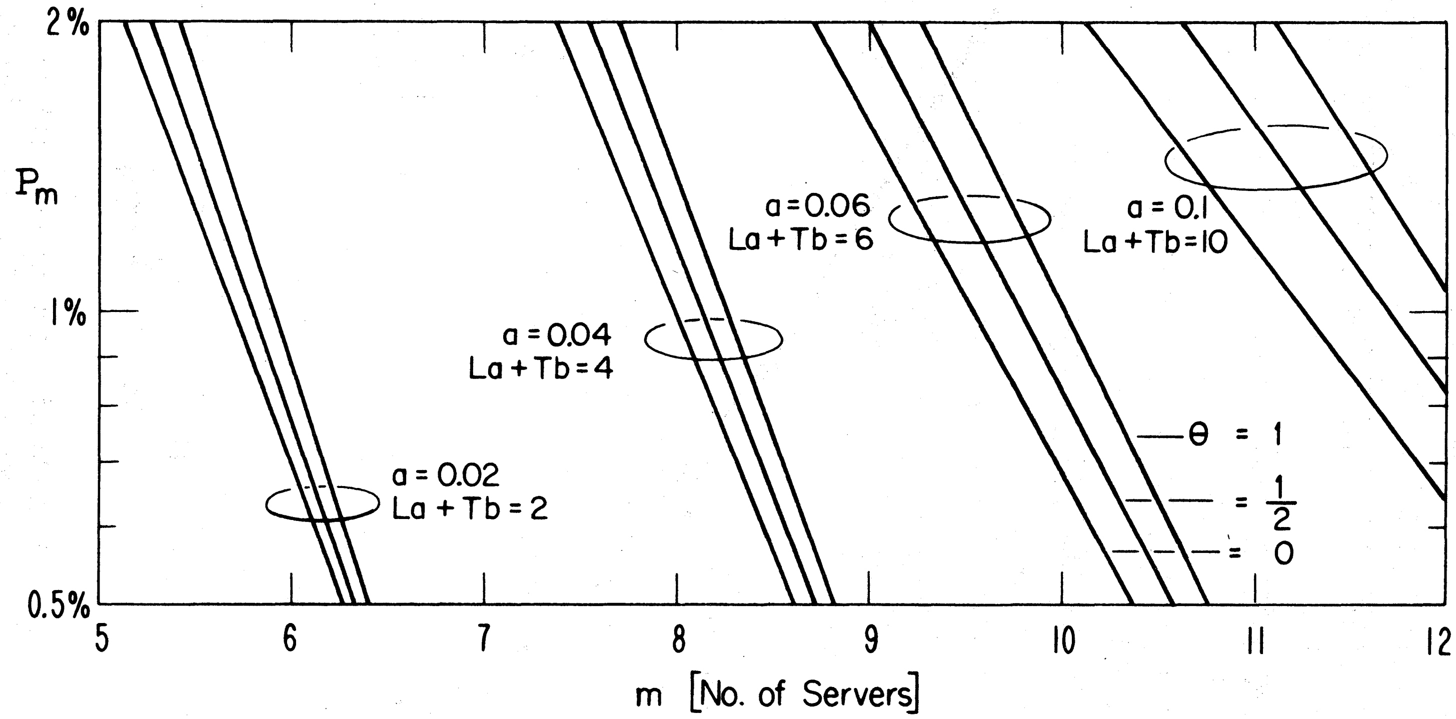


Figure B-4. Blocking probability approximations for various initial loads, La+Tb, versus the number of servers m. For L=50, T=10 and b=5a.

sufficiently large number of loads, as well as for the corresponding choices of numbers of servers, the curves of Figure 45 (main text) materialize.

B.3. References

Abramowitz, M., and I.A. Stegun, Eds. (1964), Handbook of Mathematical Functions, National Bureau of Standards, Applied Math. Series (55), Washington, DC.

AT&T (1961), Switching Systems, American Telephone and Telegraph Co., New York, NY.

Cooper, R.B. (1972), Introduction to Queueing Theory (MacMillan Co., New York, NY).

Kleinrock, L. (1975), Queueing Systems, Volume I: Theory (John Wiley and Sons, New York, NY).

Nesenbergs, M. (1979), A hybrid of Erlang B and C formulas and its applications, IEEE Trans. on Comm., COM-27, January, pp. 59-68.

Siemens (1970), Telephone Traffic Theory Tables and Charts (Siemens AG, Munchen, Germany).

APPENDIX C. THREE KINDS OF CALLS AT A LINE/TRUNK SWITCH

C.1. Assumptions

In this section, the problem of three kinds of calls through a switching network is examined further. This is the same problem that was introduced in Section 5.4 of the main text. An illustration of the model is given in Figure 46. One assumes:

- o Blocked calls are lost (cleared).
- o Arrivals are generated by three independent Poisson processes. The interarrival times are exponential.
- o The service or holding times are exponentially distributed with the same unity mean for all.
- o The number of lines (L) and trunks (T) are given constants.
- o All three service requests come from infinite user populations.
- o The switching network is nonblocking.
- o The switching network provides full availability.
- o The offered loads (arrival rates) are different for the three service classes.

To be more specific, assume the notation summarized in Table C-1. When ℓ local calls, d distant calls, and t tandem calls are in progress, one says that the system is in state (ℓ,d,t). Individual states may occur with different frequencies. Only states with $\min(\ell,d,t)\geq 0$ and $\max(2\ell+d-L, d+2t-T)\leq 0$ can ever occur. As before, the steady state or equilibrium probability of a state (ℓ,d,t) shall be denoted by $p(\ell,d,t)$. Impossible states (ℓ,d,t) will have $p(\ell,d,t)=0$ by definition. Selected sets of $p(\ell,d,t)>0$ contribute to the three sought blocking probabilities: P_ℓ for local calls, P_d for distant calls, and P_t for tandem calls.

Table C-1. Summary and Notation for the Three Kinds of Traffic

Kind of Call	State Index	Prob. of Blocking	Offered Load	Used Number of	
				Lines	Trunks
Local	ℓ	P_ℓ	A	2	0
Distant	d	P_d	B	1	1
Tandem	t	P_t	C	0	2

Let the offered loads be A, B, and C Erlangs for the local, distant and tandem calls, respectively. Finally, note that a local call cannot be completed unless there are at least two idle lines available. Likewise, a distant call demands at least one idle line plus one idle trunk. A tandem call needs two trunks. Otherwise blocking occurs. Table C-2 lists all the blocking eventualities as a function of the state descriptors (ℓ,d,t). For example, if $2\ell+d=L-1$ and $d+2t=T-1$, then only local and tandem calls are blocked. An additional distant call may still be accepted. In the lower right corner of each cell there is a number $i=1,2,\ldots,9$. This number is used solely for indexing or subscripting. For our previous example of $2\ell+d=L-1$ and $d+2t=T-1$, $i=5$.

Table C-2. Occurrence of Local, Distant, and Tandem Call Blocking as a Function of State Index Set (ℓ,d,t)

	$d+2t=T$	$d+2t=T-1$	$d+2t\leq T-2$
$2\ell+d=L$	Local Distant Tandem 1	Local Distant Tandem 2	Local Distant -- 4
$2\ell+d=L-1$	Local Distant Tandem 3	Local -- Tandem 5	Local -- -- 7
$2\ell+d\leq L-2$	-- Distant Tandem 6	-- -- Tandem 8	-- -- -- 9

C.2. Equilibrium Equations

This section constructs the equilibrium equations for the unknown three-dimensional state probabilities $p(\ell,d,t)$, where $\ell,d,t=0,1,2,\ldots$ Previously, similar identities were called steady-state flow conservation laws or "OUT=IN" equations [see (A-2), (A-5), and (B-1)]. In this problem, there is sufficient additional complexity to warrant a formal attempt at simplification.

To start, note the normalization property

$$\sum_{\text{all } \ell,d,t} p(\ell,d,t) = 1. \qquad \text{(C-1)}$$

Also note, as pointed out earlier, that $p(\ell,d,t)$ must vanish whenever either $\min(\ell,d,t)<0$ or when $\max(2\ell+d-L,\ d+2t-T)>0$.

In what follows, the equilibrium equations will be written in the form of vanishing vector products,

$$\underline{H} \cdot \underline{V} = 0. \qquad \text{(C-2)}$$

Here, the dot (•) denotes the ordinary inner product of two seven-dimensional vectors. The vector $\underline{V}$ will always be a column (vertical) vector,

$$\underline{V} = \begin{bmatrix} p(\ell,d,t) \\ p(\ell-1,d,t) \\ p(\ell,d-1,t) \\ p(\ell,d,t-1) \\ p(\ell+1,d,t) \\ p(\ell,d+1,t) \\ p(\ell,d,t+1) \end{bmatrix} \qquad \text{(C-3)}$$

It represents the unknowns in this problem.

The $\underline{H}$ will be a row (horizontal) vector, also with seven elements. Depending on system states, vector $\underline{H}$ will take various forms. These forms are introduced next with the aid of constituent vectors $\underline{h}_i$, $i=1,2,\ldots,5$:

$$\begin{aligned}
\underline{h}_1 &= [1\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ 0], \\
\underline{h}_2 &= [\ell+d+t,\ -A,\ -B,\ -C,\ \ 0\ ,\ 0\ ,\ 0], \\
\underline{h}_3 &= [0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ -(\ell+1),\ 0\ ,\ 0], \\
\underline{h}_4 &= [0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ -(d+1),\ 0], \\
\underline{h}_5 &= [0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ 0\ ,\ -(t+1)].
\end{aligned} \qquad \text{(C-4)}$$

Next, for each of the nine state categories of Table C-2, the row vector $\underline{H}_j (j=1,\ldots,9)$ is constructed as follows:

$$\begin{aligned}
\underline{H}_1 &= \underline{H}_2 = \underline{H}_3 = \underline{h}_2, \\
\underline{H}_4 &= C\underline{h}_1 + \underline{h}_2, \\
\underline{H}_5 &= B\underline{h}_1 + \underline{h}_2 + \underline{h}_4, \\
\underline{H}_6 &= A\underline{h}_1 + \underline{h}_2 + \underline{h}_3, \\
\underline{H}_7 &= (B+C)\underline{h}_1 + \underline{h}_2 + \underline{h}_4 + \underline{h}_5, \\
\underline{H}_8 &= (A+B)\underline{h}_1 + \underline{h}_2 + \underline{h}_3 + \underline{h}_4, \\
\underline{H}_9 &= (A+B+C)\underline{h}_1 + \underline{h}_2 + \underline{h}_3 + \underline{h}_4 + \underline{h}_5.
\end{aligned} \qquad \text{(C-5)}$$

Returning to (C-2), one now asserts that

$$\underline{H}_j \cdot \underline{V} = 0, \qquad j = 1,\ldots,9, \qquad \text{(C-6)}$$

represents the equilibrium equation for the state j, as indexed in each cell of Table A-2.

For instance, when $2\ell+d=L-1$ and $d+2t=T-1$ hold, one has $j=5$. For this case $\underline{H}_5 \cdot \underline{V}=0$ boils down to

$$(B+\ell+d+t)p(\ell,d,t) - Ap(\ell-1,d,t) - Bp(\ell,d-1,t) - Cp(\ell,d,t-1) - (d+1)p(\ell,d+1,t) = 0. \quad \text{(C-7)}$$

The complete equation set to be solved consists of (C-3) to (C-6). One finds Table C-2 helpful to verify that no terms are overlooked in this complex construction.

Assume that $T \le L$ and that both T and L are even. Then the number of unknowns, which is the same as the number of equations, can be found. For each d in the set, $d=0,1,2,\ldots, T/2$, the summation over all possible ℓ and t is relatively simple. The resultant number of unknowns equals

$$\frac{T+2}{4}\left[T\left(\frac{T+1}{3} + \frac{L-T}{2}\right) + \frac{L+2}{4}\right] \cong \frac{LT^2}{8}. \quad \text{(C-8)}$$

For a modestly sized switching facility, such as given by L=140 and T=24, the number of equations (same as the number of unknowns) amounts to roughly 10,000.

C.3. The Formal Solution

Given the above problem of considerable complexity, it seems rather fortunate that a solution can be constructed by a simple method. That method is the same separability device or product method (Cooper, 1972) employed in Appendix B [see, for instance, equations (B-2) and (B-3)].

Assume

$$p(\ell,d,t) = (\text{constant})p_1(\ell)p_2(d)p_3(t). \quad \text{(C-9)}$$

Then the equilibrium equations (C-6) break down to separate equations for $p_1(\ell)$, $p_2(d)$, and $p_3(t)$ that can be solved by inspection. The result is

$$p(\ell,d,t) = p(0,0,0)\,\frac{A^\ell B^d C^t}{\ell!\,d!\,t!}, \quad \text{(C-10)}$$

where

$$p(0,0,0) = \left[\sum_{\text{all } i,j,k} \frac{A^i B^j C^k}{i!\,j!\,k!}\right]^{-1}. \quad \text{(C-11)}$$

It is beyond the scope of this short appendix to prove that every condition is satisfied by this solution. However, a short example may suffice.

Consider (C-6) for j=9. With the aid of (C-3), (C-4), (C-5) and (C-9), this reduces to

$$\begin{aligned}&[(A+\ell)p_1(\ell) - Ap_1(\ell-1) - (\ell+1)p_1(\ell+1)]/p_1(\ell)\\ &+ [(B+d)p_2(d) - Bp_2(d-1) - (d+1)p_2(d+1)]/p_2(d)\\ &+ [(C+t)p_3(t) - Cp_3(t-1) - (t+1)p_3(t+1)]/p_3(t) = 0.\end{aligned} \tag{C-12}$$

Clearly, if all the square brackets vanish, one has a solution. But the expressions in the square brackets are akin to those that occur for the lost calls held, or the $M/M/\infty$, model that is known to result in the Poisson formula (Riordan, 1962; Kleinrock, 1975; Bear, 1976). Thus

$$\begin{aligned}p_1(\ell) &= p_1(0)A^{\ell}/\ell!,\\ p_2(d) &= p_2(0)B^{d}/d!,\\ p_3(t) &= p_3(0)C^{t}/t!.\end{aligned} \tag{C-13}$$

This formal solution is likewise valid for all other j=1,...,8. The constant in (C-9) is easily determined with the help of the normalization (C-1). It leads to the p(0,0,0) expression given in (C-11).

The blocking probabilities P_ℓ, P_d and P_t (see Tables C-1 and C-2) are sums of $p(\ell,d,t)$'s over appropriate (ℓ,d,t) sets. For example, the index set for local call blocking is (1,2,3,4,5,7). For distant calls blocking occurs over (1,2,3,4,6), and for tandem calls in (1,2,3,5,6,8). Their numerical evaluation seems to call for thorough planning, followed by extensive use of computing machines.

C.4. References

Bear, D. (1976), Principles of Telecommunication - Traffic Engineering (P. Peregrinus Ltd., on behalf of IEE, Telecomm. Series 2, Stevenage, England).

Cooper, R.B. (1972), Introduction to Queueing Theory (MacMillan Co., New York, NY).

Kleinrock, L. (1975), Queueing Systems, Volume I: Theory (John Wiley and Sons, New York, NY).

Riordan, J. (1962), Stocastic Service Systems (John Wiley and Sons, New York, NY).

FORM OT-29
(3-73)

U.S. DEPARTMENT OF COMMERCE
OFFICE OF TELECOMMUNICATIONS

BIBLIOGRAPHIC DATA SHEET

	1. PUBLICATION OR REPORT NO.	2. Gov't Accession No.	3. Recipient's Accession No.
	NTIA Report 79-26		

4. TITLE AND SUBTITLE SWITCH ELEMENT CAPACITIES IN ACCESS AREA DIGITAL SWITCHING SYSTEMS	5. Publication Date September 1979 6. Performing Organization Code ITS
7. AUTHOR(S) R.F. Linfield and M. Nesenbergs	9. Project/Task/Work Unit No.
8. PERFORMING ORGANIZATION NAME AND ADDRESS U.S. Department of Commerce National Telecommunications & Information Administration Institute for Telecommunication Sciences 325 Broadway, Boulder, CO 80303	10. Contract/Grant No.
11. Sponsoring Organization Name and Address U.S. Army Communications Systems Agency ATTN: CCM-RD Ft. Monmouth, NJ 07703	12. Type of Report and Period Covered 13.

14. SUPPLEMENTARY NOTES

15. ABSTRACT *(A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)*

The capacity of a digital circuit switching system is defined in terms of its four major elements: the traffic offered by the interface element, the maximum traffic carried by the switch matrix, the maximum number of call attempts handled by the control processor, and the maximum number of calls handled by the signaling elements. These element capacities may be engineered for the expected traffic (e.g., call attempts and holding times), the performance objectives (e.g., blocking probabilities and tolerable delays), the processor capabilities (e.g., speeds, memory sizes, service features), and the signaling techniques (e.g., common channel or per-channel signaling).

In a properly engineered system, the interrelations between the capacity of all elements should be considered. This report discusses such interrelationships and characterizes representative switch configurations on the basis of the four major elements. The results obtained have applications in developing non-tactical networks for military access areas where the communications profile is known. This is demonstrated for Ft. Monmouth and its environs. Estimates of traffic statistics and switching requirements are made with the aid of available Ft. Monmouth terminal density profiles.

16. Key Words *(Alphabetical order, separated by semicolons)*

17. AVAILABILITY STATEMENT	18. Security Class *(This report)*	20. Number of pages
[X] UNLIMITED.	Unclassified	192
[ ] FOR OFFICIAL DISTRIBUTION.	19. Security Class *(This page)* Unclassified	21. Price:

★ U. S. GOVERNMENT PRINTING OFFICE 1979 - 681-676/105 Reg 8

USCOMM-DC 29716-P73