1

BASIC TELEPHONY

1 DEFINITION AND CONCEPT

Telecommunication deals with the service of providing electrical communication at a distance. The service is supported by an industry that depends on a large body of increasingly specialized scientists, engineers, and craftspeople. The service may be private or open to public correspondence (i.e., access). Examples of the latter are government-owned telephone companies, often called *administrations* or private corporations, that sell their services publicly.

1.1 Telecommunication Networks

The public switched telecommunication network (PSTN) is immense. It consists of hundreds of smaller networks interconnected. There are "fixed" and "mobile" counterparts. They may or may not have common ownership. In certain areas of the world the wired and *wireless* portions of the network compete. One may also serve as a backup for the other upon failure. It is estimated that by 2005 there will be as many wireless telephones as wired telephones, about 5×10^9 handsets worldwide of each variety.

These networks, whether mobile or fixed, have traditionally been based on speech operations. Meanwhile, another network type has lately gained great importance in the scheme of things. This is the *enterprise network*. Such a network supports the business enterprise. It can just as well support the government "enterprise" as a private business. Its most common configuration is a *local area network* (LAN) and is optimized for data communications, The enterprise network also has a long-distance counterpart, called a *WAN* or wide area network. The U.S. Department of Defense developed a special breed of WAN where the original concept was for resource sharing among U.S. and allied universities. Since its inception around 1987, it has taken on a very large life of its

Telecommunication System Engineering, by Roger L. Freeman
ISBN 0-471-45133-9

own, having been opened to the public worldwide. It is the *internet*. Its appeal is universal, serving its original intent as a resource-sharing medium extending way beyond the boundaries of universities and now including a universal messaging service called *email* (electronic mail).

Some may argue that telecommunications with all its possible facets is the world's largest business. We do not take sides on this issue. What we do wish to do is to impart to the reader a technical knowledge and appreciation of telecommunication networks from a system viewpoint. By *system* we mean how one discipline can interact with another to reach a certain end objective. If we do it right, that interaction will be synergistic and will work for us; if not, it may work against us in reaching our goal.

Therefore, a primary concern of this book is to describe the development of the PSTN and enterprise network and discuss why they are built the way they are and how they are evolving. The basic underpinning of the industry was telephone service. That has now changed. The greater portion of the traffic carried today is data traffic, and all traffic is in a digital format of one form or another. We include wireless/cellular and "broadband" as adjuncts of the PSTN.

Telecommunication engineering has traditionally been broken down into two basic segments: transmission and switching. This division was most apparent in conventional telephony. Transmission deals with the delivery of a quality electrical signal from point X to point Y. Let us say that switching connects X to Y, rather than to Z. When the first edition of this book was published, transmission and switching were two very distinct disciplines. Today, that distinction has disappeared, particularly in the enterprise network. As we proceed through the development of this text, we must deal with both disciplines and show in later chapters how the dividing line separating them has completely disappeared.

2 THE SIMPLE TELEPHONE CONNECTION

The common telephone as we know it today is a device connected to the outside world by a pair of wires. It consists of a handset and its cradle with a signaling device, consisting of either a dial or push buttons. The handset is made up of two electroacoustic transducers, the earpiece or receiver and the mouthpiece or transmitter. There is also a sidetone circuit that allows some of the transmitted energy to be fed back to the receiver.

The transmitter or mouthpiece converts acoustic energy into electric energy by means of a carbon granule transmitter. The transmitter requires a direct-current (dc) potential, usually on the order of 3–5 V, across its electrodes. We call this the *talk battery*, and in modern telephone systems it is supplied over the line (central battery) from the switching center and has been standardized at −48 V dc. Current from the battery flows through the carbon granules or grains when the telephone is lifted from its cradle or goes "off hook."* When sound impinges

* The opposite action of "off hook" is "on hook"—that is, placing the telephone back in its cradle, thereby terminating a connection.

on the diaphragm of the transmitter, variations of air pressure are transferred to the carbon, and the resistance of the electrical path through the carbon changes in proportion to the pressure. A pulsating direct current results.

The typical receiver consists of a diaphragm of magnetic material, often soft iron alloy, placed in a steady magnetic field supplied by a permanent magnet, and a varying magnetic field caused by voice currents flowing through the voice coils. Such voice currents are alternating (ac) in nature and originate at the far-end telephone transmitter. These currents cause the magnetic field of the receiver to alternately increase and decrease, making the diaphragm move and respond to the variations. Thus an acoustic pressure wave is set up, more or less exactly reproducing the original sound wave from the distant telephone transmitter. The telephone receiver, as a converter of electrical energy to acoustic energy, has a comparatively low efficiency, on the order of 2–3%.

Sidetone is the sound of the talker's voice heard in his (or her) own receiver. Sidetone level must be controlled. When the level is high, the natural human reaction is for the talker to lower his or her voice. Thus by regulating sidetone, talker levels can be regulated. If too much sidetone is fed back to the receiver, the output level of the transmitter is reduced as a result of the talker lowering his or her voice, thereby reducing the level (voice volume) at the distant receiver and deteriorating performance.

To develop our discussion, let us connect two telephone handsets by a pair of wires, and at middistance between the handsets a battery is connected to provide that all-important talk battery. Such a connection is shown diagrammatically in Figure 1.1. Distance D is the overall separation of the two handsets and is the sum of distances d_1 and d_2; d_1 and d_2 are the distances from each handset to the central battery supply. The exercise is to extend the distance D to determine limiting factors given a fixed battery voltage, say, 48 V dc. We find that there are two limiting factors to the extension of the wire pair between the handsets. These are the IR drop, limiting the voltage across the handset transmitter, and the attenuation. For 19-gauge wire, the limiting distance is about 30 km, depending on the efficiency of the handsets. If the limiting characteristic is attenuation and we desire to extend the pair farther, amplifiers could be used in the line. If the battery voltage is limiting, then the battery voltage could be increased. With the telephone system depicted in Figure 1.1, only two people can communicate. As soon as we add a third person, some difficulties begin to arise. The simplest approach would be to provide each person with two handsets. Thus party A

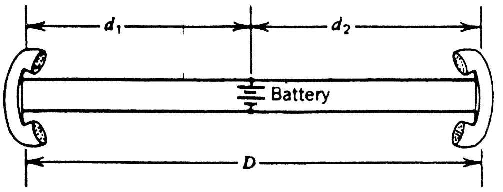


Figure 1.1. A simple telephone connection.

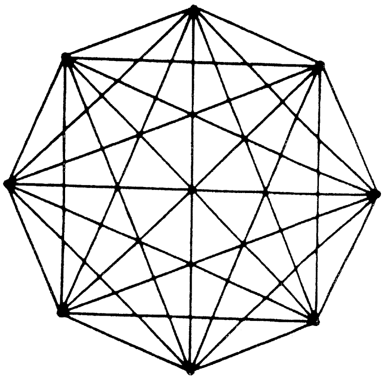


Figure 1.2. An 8-point mesh connection.

would have one set to talk to B, another to talk to C, and so forth. Or the sets could be hooked up in parallel. Now suppose A wants to talk to C and doesn't wish to bother B. Then A must have some method of selectively alerting C. As stations are added to the system, the alerting problem becomes quite complex. Of course, the proper name for this selection and alerting is *signaling*. If we allow that the pair of wires through which current flows is a loop, we are dealing with loops. Let us also call the holder of a telephone station a *subscriber*. The loops connecting them are subscriber loops.

Let us now look at an eight-subscriber system, each subscriber connected directly to every other subscriber. This is shown in Figure 1.2. When we connect each and every station with every other one in the system, this is called a *mesh* connection, or sometimes full mesh. Without the use of amplifiers and with 19-gauge copper wire size, the limiting distance is 30 km. Thus any connecting segment of the octagon may be no greater than 30 km. The only way we can justify a mesh connection of subscribers economically is when each and every subscriber wishes to communicate with every other subscriber in the network for virtually the entire day (full period). As we know, however, most telephone subscribers do not use their telephones on a full-time basis. The telephone is used at what appear to be random intervals throughout the day. Furthermore, the ordinary subscriber or telephone user will normally talk to only one other subscriber at a time. He/she will not need to talk to all other subscribers simultaneously.

If more subscribers are added and the network is extended beyond about 30 km, it is obvious that transmission costs will spiral, because that is what we are dealing with exclusively here—transmission. We are connecting each and every subscriber together with wire transmission means, requiring many amplifiers and talk batteries. Thus it would seem wiser to share these facilities in some way and cut down on the transmission costs. We now discuss this when switch and switching enter the picture. Let us define a *switch* as a device that connects inlets to outlets. The inlet may be a calling subscriber line, and the outlet may be the line of a called subscriber. The techniques of switching and the switch

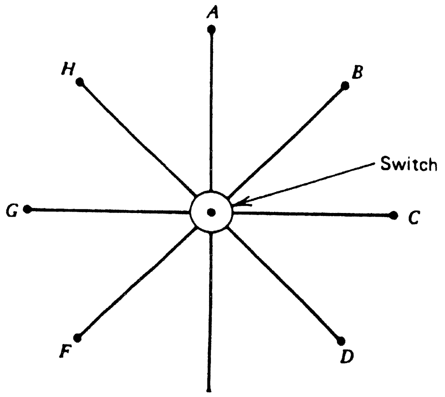


Figure 1.3. Subscribers connected in a star arrangement.

as a concept are widely discussed later in this text. Switching devices and how they work are covered in Chapters 3 and 9. Consider Figure 1.3, which shows our subscribers connected in a *star* network with a switch at the center. All the switch really does in this case is to reduce the transmission cost outlay. Actually, this switch reduces the number of links between subscribers, which really is a form of concentration. Later in our discussion it becomes evident that switching is used to concentrate traffic, thus reducing the cost of transmission facilities.

3 SOURCES AND SINKS*

Traffic is a term that quantifies usage. A subscriber *uses* the telephone when he/she wishes to talk to somebody. We can make the same statement for a telex (teleprinter service) subscriber or a data-service subscriber. But let us stay with the telephone.

A network is a means of connecting subscribers. We have seen two simple network configurations, the mesh and star connections, in Figures 1.2 and 1.3. When talking about networks, we often talk of sources and sinks. A call is initiated at a traffic source and received at a traffic sink. Nodal points or nodes in a network are the switches.

4 TELEPHONE NETWORKS: INTRODUCTORY TERMINOLOGY

From our discussion we can say that a telephone network can be regarded as a systematic development of interconnecting transmission media arranged so that one telephone user can talk to any other within that network. The evolving layout of the network is primarily a function of economics. For example, subscribers share common transmission facilities; switches permit this sharing by concentration.

* The traffic engineer may wish to use the terminology “origins and destinations.”

Consider a very simplified example. Two towns are separated by, say, 20 miles, and each town has 100 telephone subscribers. Logically, most of the telephone activity (the traffic) will be among the subscribers of the first town and among those of the second town. There will be some traffic, but considerably less, from one town to the other. In this example let each town have its own switch. With the fairly low traffic volume from one town to the other, perhaps only six lines would be required to interconnect the switch of the first town to that of the second. If no more than six people want to talk simultaneously between the two towns, a number as low as six can be selected. Economics has mandated that we install the minimum number of connecting telephone lines from the first town to the second to serve the calling needs between the two towns. The telephone lines connecting one telephone switch or exchange with another are called *trunks* in North America and *junctions* in Europe. The telephone lines connecting a subscriber to the switch or exchange that serves the subscriber are called *lines, subscriber lines*, or *loops*. Concentration is a line-to-trunk ratio. In the simple case above, it was 100 lines to six trunks (or junctions), or about a 16 : 1 ratio.

A telephone subscriber looking into the network is served by a *local exchange*. This means that the subscriber's telephone line is connected to the network via the local exchange or central office, in North American parlance. A local exchange has a serving area, which is the geographical area in which the exchange is located; all subscribers in that area are served by that exchange.

The term *local area*, as opposed to *toll area*, is that geographical area containing a number of local exchanges and inside which any subscriber can call any other subscriber without incurring tolls (extra charges for a call). Toll calls and long-distance calls are synonymous. For instance, a local call in North America, where telephones have detailed billing, shows up on the bill as a time-metered call or is covered by a flat monthly rate. Toll calls in North America appear as separate detailed entries on the telephone bill. This is not so in most European countries and in those countries following European practice. In these countries there is no detailed billing on direct-distance-dialed (subscriber-trunk-dialed) calls. All such subscriber-dialed calls, even international ones, are just metered, and the subscriber pays for the meter steps used per billing period, which is often one or two months. In European practice a long-distance call, a toll call if you will, is one involving the dialing of additional digits (e.g., more than six or seven digits).

Let us call a network a *grouping of interworking telephone exchanges*. As the discussion proceeds, the differences between local networks and national networks are shown. Two other types of network are also discussed. These are specialized versions of a local network and are the rural network (rural area) and metropolitan network (metropolitan area). (Also consult Refs. 9 and 16–18.)

5 ESSENTIALS OF TRAFFIC ENGINEERING

5.1 Introduction and Terminology

As we have already mentioned, telephone exchanges are connected by trunks or junctions. The number of trunks connecting exchange X with exchange Y is the number of voice pairs or their equivalent used in the connection. One of the most important steps in telecommunication engineering practice is to determine the number of trunks required on a route or connection between exchanges. We could say we are *dimensioning* the route. To dimension a route correctly, we must have some idea of its usage—that is, how many people will wish to talk at once over the route. The usage of a transmission route or a switch brings us into the realm of traffic engineering, and the usage may be defined by two parameters: (1) *calling rate*, or the number of times a route or traffic path is used per unit period, or, more properly defined, "the call intensity per traffic path during the busy hour";* and (2) *holding time*, or "the duration of occupancy of a traffic path by a call,"* or sometimes, "the average duration of occupancy of one or more paths by calls."* A *traffic path* is "a channel, time slot, frequency band, line, trunk, switch, or circuit over which individual communications pass in sequence."* *Carried traffic* is the volume of traffic actually carried by a switch, and *offered traffic* is the volume of traffic offered to a switch.

To dimension a traffic path or size a telephone exchange, we must know the traffic intensity representative of the normal busy season. There are weekly and daily variations in traffic within the busy season. Traffic is very random in nature. However, there is a certain consistency we can look for. For one thing, there usually is more traffic on Mondays and Fridays and a lower volume on Wednesdays. A certain consistency can also be found in the normal workday hourly variation. Across the typical day the variation is such that a 1-h period shows greater usage than any other. From the hour with least traffic to the hour of greatest traffic, the variation can exceed 100 : 1. Figure 1.4 shows a typical hour-by-hour traffic variation for a serving switch in the United States.† It can be seen that the busiest period, the *busy hour* (BH), is between 10 A.M. and 11 A.M. From one workday to the next, originating BH calls can vary as much as 25%. To these fairly "regular" variations, there are also unpredictable peaks caused by stock market or money market activity, weather, natural disaster, international events, sporting events, and so on. Normal system growth must also be taken into account. Nevertheless, suitable forecasts of BH traffic can be made. However, before proceeding, consider the five most common definitions of BH:

Busy Hour Definitions (CCITT Rec. E.600)

1. *Busy Hour.* The busy hour refers to the traffic volume or number of call attempts, and is that continuous 1-h period lying wholly in the time interval concerned for which this quantity (i.e., traffic volume or call attempts) is greatest.
2. *Peak Busy Hour.* The busy hour each day; it usually is not the same over a number of days.

* Reference Data for Radio Engineers [1], pages 31–38.

† The busy hour will vary from country to country because of cultural differences.

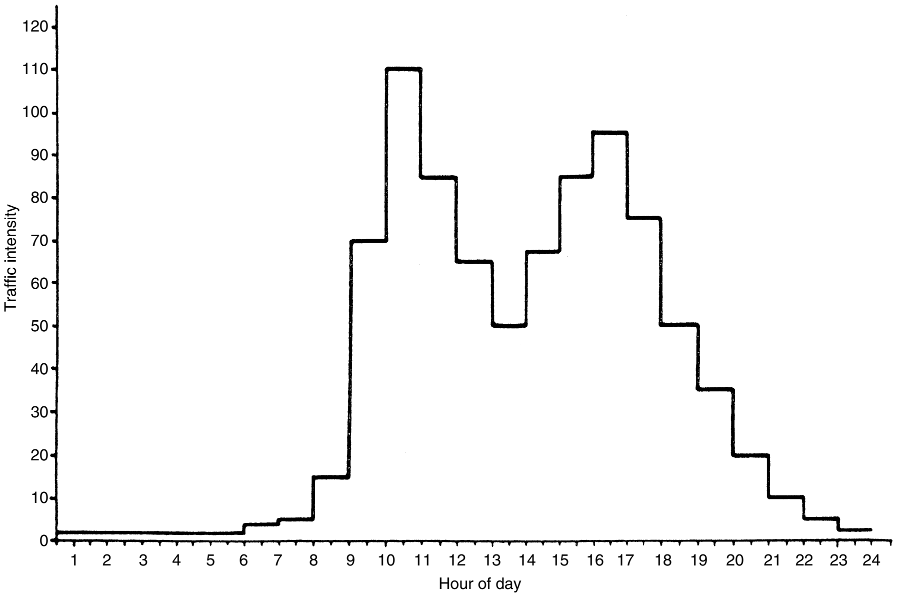


Figure 1.4. Bar chart of traffic intensity over a typical working day (United States, mixed business and residential).

3. *Time Consistent Busy Hour.* The 1-h period starting at the same time each day for which the average traffic volume or call-attempt count of the exchange or resource group concerned is greatest over the days under consideration.

From *Engineering and Operations in the Bell System*, 2nd ed. [23]

4. The engineering period (where the grade of service criteria is applied) is defined as the busy season busy hour (BSBH), which is the busiest clock hour of the busiest weeks of the year.
5. The average busy season busy hour (ABSBH) is used for trunk groups and always has a grade of service criterion applied. For example, for the ABSBH load, a call requiring a circuit in a trunk group should encounter "all trunks busy" (ATB) no more than 1% of the time.

Reference 23 goes on to state that peak loads are of more concern than average loads when engineering switching equipment and engineering periods other than the ABSBH are defined. Examples of these are the highest BSBH and the average of the ten highest BSBHs. Sometimes the engineering period is the weekly peak hour (which may not even be the BSBH).

When dimensioning telephone exchanges and transmission routes, we shall be working with BH traffic levels and care must be used in the definition of busy hour.

5.2 Measurement of Telephone Traffic

If we define *telephone traffic* as the aggregate of telephone calls over a group of circuits or trunks with regard to the duration of calls as well as their number [2], we can say that traffic flow (A) is expressed as

$$A = C \times T$$

where C designates the number of calls originated during a period of 1 h and T is the average holding time, usually given in hours. A is a dimensionless unit because we are multiplying calls/hour by hour/call.

Suppose that the average holding time is 2.5 min and the calling rate in the BH for a particular day is 237. The traffic flow (A) would then be 237×2.5, or 592.5 call-minutes (Cm) or 592.5/60, or about 9.87 call-hours (Ch).

Ramses Mina [2] states that a distinction should be made between the terms "traffic density" and "traffic intensity." The former represents the number of simultaneous calls at a given moment, while the latter represents the average traffic density during a 1-h period. The quantity of traffic used in the calculation for dimensioning of switches is the traffic intensity.

The preferred unit of traffic intensity is the erlang, named after the Danish mathematician A. K. Erlang [5]. The erlang is a dimensionless unit. One erlang represents a circuit occupied for 1 h. Considering a group of circuits, traffic intensity in erlangs is the number of call-seconds per second or the number of call-hours per hour. If we knew that a group of 10 circuits had a call intensity of 5 erlangs, we would expect half of the circuits to be busy at the time of measurement.

In the United States the term "unit call" (UC), or its synonymous term, "hundred call-second," abbreviated CCS,* generally is used. These terms express the sum of the number of busy circuits, provided that the busy trunks were observed once every 100 s (36 observations in 1 h) [2].

There are other traffic units. For instance: call-hour (Ch)—1 Ch is the quantity represented by one or more calls having an aggregate duration of 1 h; call-second (Cs)—1 Cs is the quantity represented by one or more calls having an aggregate duration of 1 s; traffic unit (TU), a unit of traffic intensity. One TU is the average intensity in one or more traffic paths carrying an aggregate traffic of 1 Ch in 1 h (the busy hour unless otherwise specified). 1 TU = 1 E (erlang) (numerically). The *equated busy hour call* (EBHC) is a European unit of traffic intensity. 1 EBHC is the average intensity in one or more traffic paths occupied in the BH by one 2-min call or an aggregate duration of 2 min. Thus we can relate our terms as follows:

$$1 \text{ erlang} = 30 \text{ EBHC} = 36 \text{ CCS} = 60 \text{ Cm}$$

assuming a 1-h time-unit interval.

* The first letter C in CCS stands for the Roman numeral 100.

Traffic measurements used for long-term network planning are usually based on the traffic in the busy hour (BH), which is usually determined based on observations and studies.

The traditional traffic measurements on trunks during a measurement interval are:

- Peg count*—calls offered
- Usage—traffic (CCS or erlangs) carried
- Overflow—call encountering all trunks busy

From these measurements, the blocking probability and mean traffic load carried by the trunk group can be calculated.

Extensive traffic measurements are made on switching systems because of their numerous traffic sensitive components. Usual measurements for a component such as a service circuit include calls carried, peg count, and usage. The typical holding time for a common-control element in a switch is considerably shorter than that for a trunk, and short sampling intervals (e.g., 10 s) or continuous monitoring are used to measure usage.

Traffic measurements for short-term network management purposes are usually concerned with detecting network congestion. Calls offered, peg count, and overflow count can be used to calculate attempts per circuit per hour (ACH) and connections per circuit per hour (CCH), with these measurements being calculated over very short time periods (e.g., 10-min intervals).

Under normal circumstances, ACH and CCH are approximately equal. Examples of abnormal conditions are:

- ACH high, CCH normal—heavy demand, excessive blockage, normal holding times for connected calls indicating that most calls switched are completed, heavy traffic but low congestion [25].
- ACH high, CCH high—heavy traffic, short trunk holding times indicate uncompleted call attempts being switched, congestion [25]. (Consult Ref. 24. Also see Refs. 7, 10, and 19.)

5.3 Blockage, Lost Calls, and Grade of Service

Assume that an isolated telephone exchange serves 5000 subscribers and that no more than 10% of the subscribers wish service simultaneously. Therefore, the exchange is dimensioned with sufficient equipment to complete 500 simultaneous connections. Each connection would be, of course, between any two of the 5000 subscribers. Now let subscriber 501 attempt to originate a call. He/she cannot

* A term taken from telephony in the older days where manual switching was prevalent. A peg board was installed by the telephone operator to keep count of offered calls. The present definition is taken from Ref. 23. "A count of all calls offered to a trunk group, usually measured for one hour. As applied to units of switching systems with common control, *peg count*, or *carried peg count*, means the number of calls actually handled."

because all the connecting equipment is busy, even though the line he/she wishes to reach may be idle. This call from subscriber 501 is termed a *lost call* or *blocked call*. He/she has met blockage. The probability of meeting blockage is an important parameter in traffic engineering of telecommunication systems. If congestion conditions are to be met in a telephone system, we can expect that those conditions will usually be met during the BH. A switch is engineered (dimensioned) to handle the BH load. But how well? We could, indeed, far overdimension the switch such that it could handle any sort of traffic peaks. However, that is uneconomical. So with a well-designed switch, during the busiest of BHs we may expect some moments of congestion such that additional call attempts will meet blockage. *Grade of service* expresses the probability of meeting blockage during the BH and is expressed by the letter p. A typical grade of service is $p = 0.01$. This means that an average of one call in 100 will be blocked or "lost" during the BH. Grade of service, a term in the Erlang formula, is more accurately defined as the *probability of blockage*. It is important to remember that lost calls (blocked calls) refer to calls that fail at *first* trial. We discuss attempts (at dialing) later, that is, the way blocked calls are handled.

We exemplify grade of service by the following problem. If we know that there are 354 seizures (lines connected for service) and 6 blocked calls (lost calls) during the BH, what is the grade of service?

$$\begin{aligned}\text{Grade of service} &= \frac{\text{Number of lost calls}}{\text{Total number of offered calls}}\\ &= \frac{6}{354+6} = \frac{6}{360} \qquad (1.1)\end{aligned}$$

or

$$p = 0.017$$

The average grade of service for a network may be obtained by adding the grade of service contributed by each constituent switch, switching network, or trunk group. The *Reference Data for Radio Engineers* [1, Section 31] states that the grade of service provided by a particular group of trunks or circuits of specified size and carrying a specified traffic intensity is the probability that a call offered to the group will find available trunks already occupied on first attempt. That probability depends on a number of factors, the most important of which are (1) the distribution in time and duration of offered traffic (e.g., random or periodic arrival and constant or exponentially distributed holding time), (2) the number of traffic sources [limited or high (infinite)], (3) the availability of trunks in a group to traffic sources (full or restricted availability), and (4) the manner in which lost calls are "handled."

Several new concepts are suggested in these four factors. These must be explained before continuing.

5.4 Availability

Switches were previously discussed as devices with lines and trunks, but better terms for describing a switch are “inlets” and “outlets.” When a switch has full availability, each inlet has access to any outlet. When not all the free outlets in a switching system can be reached by inlets, the switching system is referred to as one with “limited availability.” Examples of switches with limited and full availability are shown in Figures 1.5A and 1.5B.

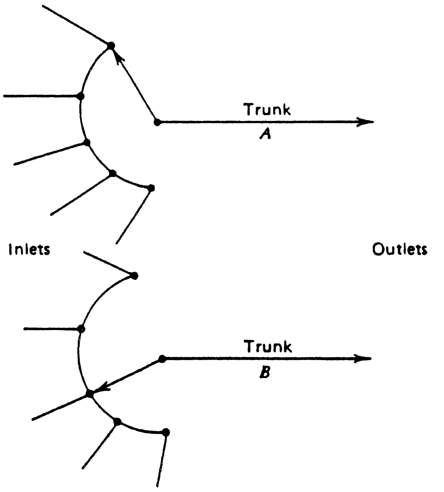


Figure 1.5A. An example of a switch with limited availability.

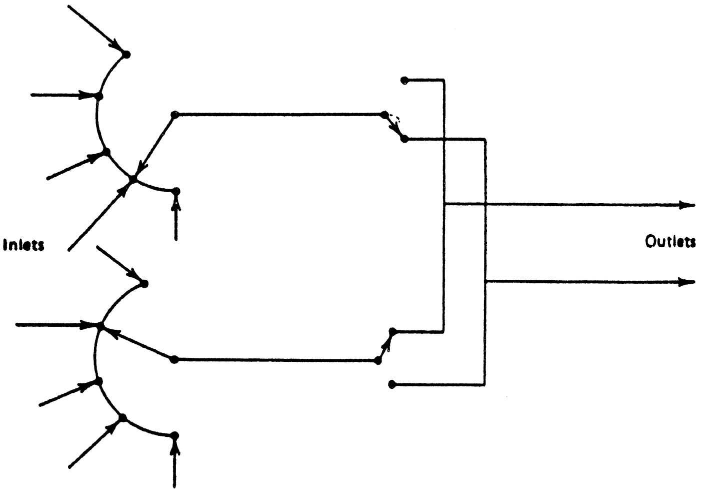


Figure 1.5B. An example of a switch with full availability.

Of course, full availability switching is more desirable than limited availability but is more expensive for larger switches. Thus full availability switching is generally found only in small switching configurations and in many new digital switches (see Chapter 9). *Grading* is one method of improving the traffic-handling capacities of switching configurations with limited availability. Grading is a scheme for interconnecting switching subgroups to make the switching load more uniform.

5.5 "Handling" of Lost Calls

In conventional telephone traffic theory, three methods are considered for the handling or dispensing of lost calls: (1) lost calls held (LCH), (2) lost calls cleared (LCC), and (3) lost calls delayed (LCD). The LCH concept assumes that the telephone user will immediately reattempt the call on receipt of a congestion signal and will continue to redial. The user hopes to seize connection equipment or a trunk as soon as switching equipment becomes available for the call to be handled. It is the assumption in the LCH concept that lost calls are held or waiting at the user's telephone. This concept further assumes that such lost calls extend the average holding time theoretically, and in this case the average holding time is zero, and all the time is waiting time. The principal traffic formula used in North America is based on the LCH concept.

The LCC concept, which is used primarily in Europe or those countries accepting European practice, assumes that the user will hang up and wait some time interval before reattempting if the user hears the congestion signal on the first attempt. Such calls, it is assumed, disappear from the system. A reattempt (after the delay) is considered as initiating a new call. The Erlang formula is based on this criterion.

The LCD concept assumes that the user is automatically put in queue (a waiting line or pool). For example, this is done when the operator is dialed. It is also done on most modern computer-controlled switching systems, generally referred to under the blanket term *stored program control* (SPC). The LCD category may be broken down into three subcategories, depending on how the queue or pools of waiting calls is handled. The waiting calls may be handled last in first out (LIFO), first in first out (FIFO), or at random.

5.6 Infinite and Finite Sources

We can assume that traffic sources are infinite or finite. For the case of infinite traffic sources, the probability of call arrival is constant and does not depend on the state of occupancy of the system. It also implies an infinite number of call arrivals, each with an infinitely small holding time. An example of finite sources is when the number of sources offering traffic to a group of trunks or circuits is comparatively small in comparison to the number of circuits. We can also say that with a finite number of sources, the arrival rate is proportional to the number of sources that are not already engaged in sending a call.

5.7 Probability-Distribution Curves

Telephone-call originations in any particular area are random in nature. We find that originating calls or call arrivals at an exchange closely fit a family of probability-distribution curves following a Poisson distribution. The Poisson distribution is fundamental to traffic theory.

Most of the common probability-distribution curves are two-parameter curves. That is, they may be described by two parameters, mean and variance. The mean is a point on the probability-distribution curve where an equal number of events occur to the right of the point and to the left of the point.

Mean is synonymous with *average*. We define mean as the x-coordinate of the center of the area under the probability-density curve for the population. The lowercase Greek letter mu (μ) is the traditional indication of the mean; $\overline{x}$ is also used.

The second parameter used to describe a distribution curve is the dispersion, which tells us how the values or population are dispersed about the center or mean of the curve. There are several measures of dispersion. One is the familiar *standard deviation*, where the standard deviation s of a sample of n observations $x_1, x_2, \ldots, x_n$ is

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{1.2}$$

The *variance* V of the sample values is the square of s. The parameters for dispersion s and s^2, the standard deviation and variance, respectively, are usually denoted σ and σ^2 and give us an idea of the squatness of a distribution curve. Mean and standard deviation of a normal distribution curve are shown in Figure 1.6, where we can see that σ^2 is another measure of dispersion, the variance, or essentially the average of the squares of the distances from mean aside from the factor $n/(n-1)$.

We have introduced two distribution functions describing the probability of distribution, often called the *distribution* of x or just $f(x)$. Both functions are used in traffic engineering. But before proceeding, the variance-to-mean ratio

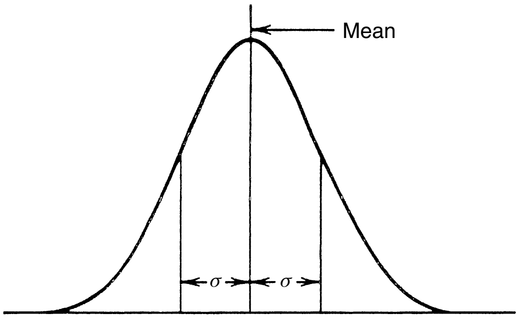


Figure 1.6. A normal distribution curve showing the mean and the standard deviation, σ.

(VMR) is introduced. Sometimes VMR(α) is called the *coefficient of overdispersion.* The formula for VMR is

$$\alpha = \frac{\sigma^2}{\mu} \tag{1.3}$$

5.8 Smooth, Rough, and Random Traffic

Traffic probability distributions can be divided into three distinct categories: (1) smooth, (2) rough, and (3) random. Each may be defined by α, the VMR. For smooth traffic, α is less than 1. For rough traffic, α is greater than 1. When α is equal to 1, the traffic distribution is called *random.* The Poisson distribution function is an example of random traffic where VMR = 1. Rough traffic tends to be peakier than random or smooth traffic. For a given grade of service, more circuits are required for rough traffic because of the greater spread of the distribution curve (greater dispersion).

Smooth traffic behaves like random traffic that has been filtered. The filter is the local exchange. The local exchange looking out at its subscribers sees call arrivals as random traffic, assuming that the exchange has not been overdimensioned. The smooth traffic is the traffic on the local exchange outlets. The filtering or limiting of the peakiness is done by call blockage during the BH. Of course, the blocked traffic may actually overflow to alternative routes. Smooth traffic is characterized by a positive binomial distribution function, perhaps better known to traffic people as the *Bernoulli distribution.* An example of the Bernoulli distribution is as follows [6]. If we assume that subscribers make calls independently of each other and that each has a probability p of being engaged in conversation, then if n subscribers are examined, the probability that x of them will be engaged is

$$\begin{aligned} B(x) &= C_x^n p^x (1-p)^{n-x}, \qquad 0 < x < n \\ \text{Its mean} &= np \\ \text{Its variance} &= np(1-p) \end{aligned} \tag{1.4}$$

where the symbol C_x^n means the number of ways that x entities can be taken n at a time. Smooth traffic is assumed in dealing with small groups of subscribers; the number 200 is often used as the breakpoint [6]. That is, groups of subscribers are considered small when the subscribers number is less than 200. And as mentioned, smooth traffic is also used with carried traffic. In this case the rough or random traffic would be the offered traffic.

Let's consider the binomial distribution for rough traffic. This is characterized by a negative index. Therefore, if the distribution parameters are k and q, where k is a positive number representing a hypothetical number of traffic sources and q represents the occupancy per source and may vary between 0 and 1, then

$$R'(x,k,q) = \binom{x+k-1}{k-1} q^x (1-q)^k \tag{1.5}$$

where R' is the probability of finding x calls in progress for the parameters k and q [2]. Rough traffic is used in dimensioning toll trunks with alternative routing. The symbol B (Bernoulli) is used by traffic engineers for smooth traffic and R for rough traffic. Although P may designate probability, in traffic engineering it designates Poissonian, and hence we have "P" tables such as those in Ref. 20, Table 1-1.

The Bernoulli formula is

$$B'(x, s, h) = C_s^x h^x (1 - h)^{s-x} \tag{1.6}$$

where C_s^x indicates the number of combinations of s things taken x at a time, h is the probability of finding the first line of an exchange busy, $1 - h$ is the probability of finding the first line idle, and s is the number of subscribers. The probability of finding two lines busy is h^2, the probability of finding s lines busy is h^s, and so on. We are interested in finding the probability of x of the s subscribers with busy lines.

The Poisson probability function can be derived from the binomial distribution, assuming that the number of subscribers s is very large and the calling rate per line h is low* such that the product $sh = m$ remains constant and letting s increase to infinity in the limit

$$P(x) = \frac{m^x}{x!} e^{-m} \tag{1.7}$$

where

$$x = 0, 1, 2, \ldots$$

For most of our future discussion, we consider call-holding times to have a negative exponential distribution in the form

$$P = e^{-t/h} \tag{1.8}$$

where t/h is the average holding time and in this case P is the probability of a call lasting longer than t, some arbitrary time interval.

Figure 1.7 compares smooth, random, and rough traffic probability distributions.

6 ERLANG AND POISSON TRAFFIC FORMULAS

When dimensioning a route, we want to find the number of circuits that serve the route. There are several formulas at our disposal to determine that number of circuits based on the BH traffic load. In Section 5.3 four factors were discussed that will help us to determine which traffic formula to use given a particular set of circumstances. These factors primarily dealt with (1) call arrivals and holding-time distribution, (2) number of traffic sources, (3) availability, and (4) handling of lost calls.

* For example, less than 50 millierlangs (mE).

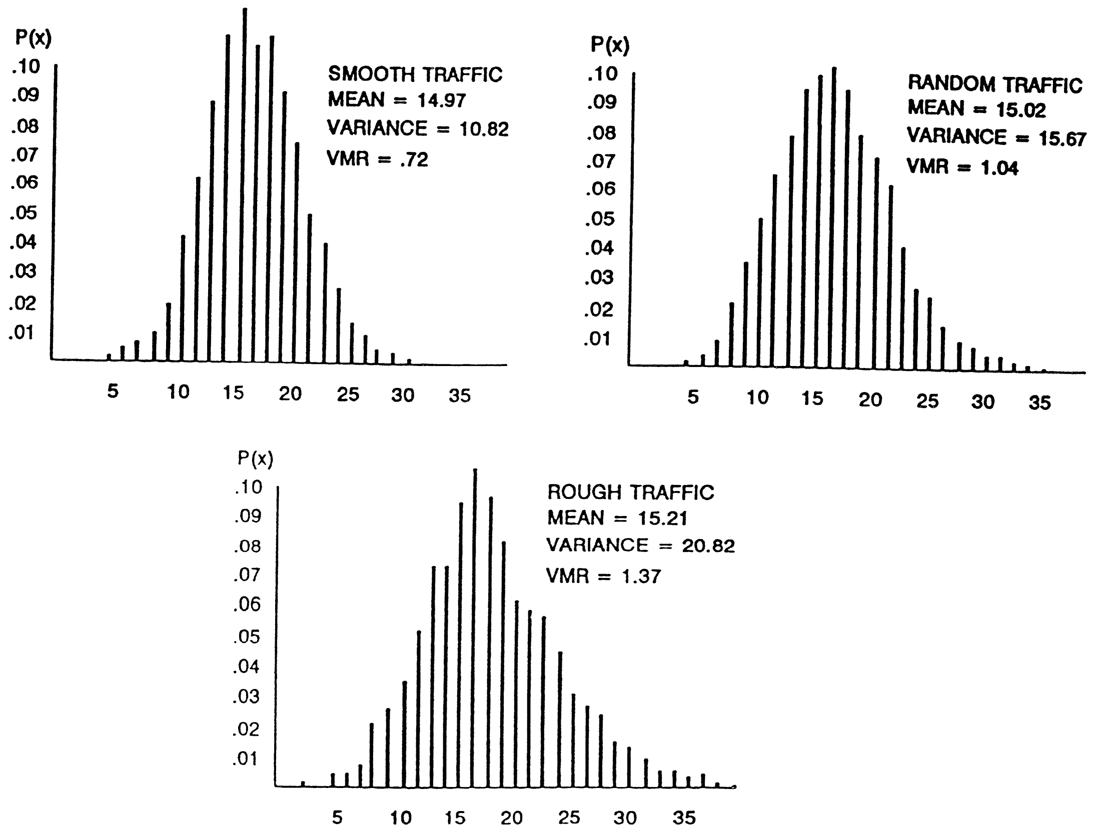


Figure 1.7. Traffic probability distributions: smooth, random, and rough traffic. Courtesy of John Lawlor Associates, Sharon, Massachusetts [25].

The Erlang B loss formula has been widely used today outside of the United States. Loss here means the probability of blockage at the switch due to congestion or to "all trunks busy" (ATB). This is expressed as *grade of service* (E_B) or the probability of finding x channels busy. The other two factors in the Erlang B formula are the mean of the *offered* traffic and the number of trunks of servicing channels available. Thus

$$E_B = \frac{A^n/n!}{1 + A + A^2/2! + \cdots + A^n/n!} \tag{1.9}$$

where n is the number of trunks or servicing channels, A is the mean of the offered traffic, and E_B is the grade of service using the Erlang B formula. This formula assumes the following:

- Traffic originates from an infinite number of sources.
- Lost calls are cleared assuming a zero holding time.
- The number of trunks or servicing channels is limited.
- Full availability exists.

At this point in our discussion of traffic we suggest that the reader learn to differentiate between time congestion and call congestion when dealing with grade of service. *Time congestion*, of course, refers to the decimal fraction of an hour during which all trunks are busy simultaneously. *Call congestion*, on the other hand, refers to the number of calls that fail at first attempt, which we term *lost calls*. Keep in mind that the Erlang B formula deals with offered traffic, which differs from carried traffic by the number of lost calls.

Table 1-2 in Ref. 20 is based on the Erlang B formula and gives trunk dimensioning information for several specific grades of service, from 0.001 to 0.10 and from 1 to 200 trunks. The traffic intensity units are in CCS and erlangs for 1 to 200 trunks. Keep in mind that 1 erlang = 36 CCS (based on a 1-h time interval). As an example of how we might employ an Erlang B table, suppose we wished a grade of service of 0.001 and the route carried 16.68 erlangs of traffic, we would then see that 30 trunks would be required. When sizing a route for trunks or when dimensioning an exchange, we often come up with a fractional numbering of servicing channels or trunks. In this case we would opt for the next highest integer because we cannot install a fraction of a trunk. For example, if calculations show that a trunk route should have 31.4 trunks, it would be designed for 32 trunks.

The Erlang B formula is based on lost calls cleared. It is generally accepted as a standard outside the United States (see CCITT Rec. Q.87). In the United States the Poisson formula (2) is favored. The formula is often called the *Molina formula*. It is based on the LCH concept. Table 1-1 of Ref. 20 provides trunking sizes for various grades of service from 0.001 to 0.10. The units of traffic intensity are the CCS and the erlang.

Table 1-1 of Ref. 20 is based on full availability. Also, we should remember that the Poisson formula also assumes that traffic originates from a large (infinite) number of independent subscribers or sources (i.e., random traffic input) with a limited number of trunks or servicing channels and LCH.

It is not as straightforward as it may seem when comparing grades of service between Poisson and Erlang B formulas (or tables). The grade of service $p = 0.01$ for the Erlang B formula is equivalent to a grade of service of 0.005 when applying the Poisson (Molina) formula. Given these grades of service, assuming LCC with the Erlang B formula permits up to several tenths of erlangs less of traffic when dimensioning up to 22 trunks, where the two approaches equate (e.g., where each formula allows 12.6 erlangs over the 22 trunks). Above 22 trunks the Erlang B formula permits the trunks to carry somewhat more traffic, and at 100 trunks it permits 2.7 erlangs more than for the Poisson formula under the LCH assumption.

6.1 Alternative Traffic Formula Conventions

Some readers may be more comfortable using traffic formulas with a different convention and notation. The Erlang B and Poisson formulas were derived from Ref. 2. The formulas and notation used in this subsection have been taken from Ref. 20. The following is the notation used in the formulas given below:

A = The expected traffic density, expressed in busy hour erlangs.
P = The probability that calls will be lost (or delayed) because of insufficient channels.
n = The number of channels in the group of channels.
s = The number of sources in the group of sources.
p = The probability that a single source will be busy at an instant of observation. This is equal to *A/s*.
x = A variable representing a number of busy sources or busy channels.
e = The Naperian logarithmic base, which is the constant 2.71828+.
$\binom{m}{n}$ = The combination of m things taken n at a time.
$\sum_{X=m}^{n}$ = The summation of all values obtained when each integer or whole number value, from m to n inclusive, is substituted for the value x in the expression following the symbol.
∞ = The conventional symbol for infinity.

The Poisson formula has the following assumptions: (1) infinite sources, (2) equal traffic density per source, and (3) lost calls held (LCH). The formula is

$$P = e^{-A} \sum_{x=n}^{\infty} \frac{A^x}{x!} \tag{1.10}$$

The Erlang B formula assumes (1) infinite sources, (2) equal traffic density per source, and (3) lost calls cleared (LCC). The formula is

$$P = \frac{\dfrac{A^n}{n!}}{\displaystyle\sum_{x=0}^{n} \frac{A^x}{x!}} \tag{1.11}$$

The Erlang C formula, commonly used with digital switching where one would expect to find queues, assumes (1) infinite sources, (2) lost calls delayed (LCD), (3) exponential holding times, and (4) calls served in order of arrival. Refer to Table 1-3 in Section 1-5 of Ref. 20. The formula is

$$P = \frac{\dfrac{A^n}{n!} \cdot \dfrac{n}{n-A}}{\displaystyle\sum_{x=0}^{n-1} \frac{A^x}{x!} + \frac{A^n}{n!}\frac{n}{n-A}} \tag{1.12}$$

The binomial formula assumes (1) finite sources, (2) equal traffic density per source, and (3) lost calls held (LCH). The formula is

$$P = \left(\frac{s-A}{s}\right)^{s-1} \sum_{x=n}^{s-1} \binom{s-1}{x} \left(\frac{A}{s-A}\right)^{x} \tag{1.13}$$

(Also consult Ref. 19.)

6.2 Computer Programs for Traffic Calculations

6.2.1 Erlang B Computer Program. Relatively simple programs can be written in BASIC computer language to solve the Erlang B equation shown below:

$$P = \frac{A^N/N!}{1 + A + A^2/2! + A^3/3! + \cdots + A^N/N!}$$

Following is a sample program. It is written using GWBASIC, Version 3.11.

```
 10  PRINT “ERLANG B, TRAFFIC ON EACH TRUNK.”
 20  INPUT “OFFERED TRAFFIC (ERLANG) = ”, A
 30  LET B = 0: LET H = 0
 40  INPUT “MINIMUM GRADE OF SERVICE (DEFAULT = 1) = ”, H
 50  IF H = 0 THEN LET H = 1
 60  INPUT “OBJECTIVE GRADE OF SERVICE (DEFAULT = 0.001) = ”, B
 70  IF B = 0 THEN LET B = .001
 80  LET N = 0: LET T = 1: LET T1 = 1: LET 0 = A: LET Z = 15
 90  PRINT “TRUNKS OFFERED CARRIED CUMULATIVE P = G/S”
100  PRINT
```

```
"-----------------------------------------------------------------------------"
110  IF T1 > 9.999999E + 20 THEN 230
120  LET N = N + 1: LET T = T * A/N: LET T1 = T1 + T: LET P = T/T1
130  LET L = A * P: LET S = A - L: LET C = 0 - L
140  IF P > H THEN LET Z = N + 14: GOTO 110
150  PRINT USING "###            ";N,
160  PRINT USING "###.####     ";0,
170  PRINT USING "      .####     ";C,
180  PRINT USING "###.####     ";S,
190  PRINT USING ".####";P
200  LET 0 = L: IF P < B THEN 240
210  IF N < Z THEN 110
220  STOP: LET Z = Z + 15: PRINT "OFFERED TRAFFIC (ERLANG) IS ", A: GOTO 90
230  LET T = T * 1E - 37: LET T1 = T1 * 1E - 37: GOTO 120
240  END
```

For a given traffic load, this program shows the load offered to each trunk, the load carried by each trunk, the cumulative load carried by the trunk group, and the probability of blocking for a group with "N" trunks. This assumes that the trunks are used in sequential order. The results are presented on a trunk-by-trunk basis until the probability of blocking reaches some minimum value [Pr(Blocking) $= 0.001$ in this version]. Both the minimum and maximum probability values can be adjusted by entering input values when starting the program.

The variable "Z" is used to stop the calculation routine every "Z" lines so that the data will fill one screen at a time. Calculation will continue when the instruction "CONT" is entered (or by pressing the "F5" key in some computers).

In order to avoid overflow in the computer registers when very large numbers are encountered, the least significant digits are dropped for large traffic loads (see Instruction 110 and Instruction 230). The variables affected by this adjustment (T and T1) are used only to determine the probability ratio, and the error introduced is negligible. As the traffic load A becomes very large, the error increases.

Following is a sample program calculation for an offered traffic load of 10 erlangs:

```
ERLANG B, TRAFFIC ON EACH TRUNK.
OFFERED TRAFFIC (ERLANG) = 10
MINIMUM GRADE OF SERVICE (DEFAULT = 1) =
OBJECTIVE GRADE OF SERVICE (DEFAULT = 0.001) =
TRUNKS      OFFERED        CARRIED        CUMULATIVE        P = G/S
-------------------------------------------------------------------------
  1          10.0000        0.9091          0.9091           0.9091
  2           9.0909        0.8942          1.8033           0.8197
  3           8.1967        0.8761          2.6794           0.7321
  4           7.3206        0.8540          3.5334           0.6467
  5           6.4666        0.8271          4.3605           0.5640
  6           5.6395        0.7944          5.1549           0.4845
  7           4.8451        0.7547          5.9096           0.4090
  8           4.0904        0.7072          6.6168           0.3383
  9           3.3832        0.6511          7.2679           0.2732
 10           2.7321        0.5863          7.8542           0.2146
 11           2.1458        0.5135          8.3677           0.1632
```

12	1.6323	0.4349	8.8026	0.1197
13	1.1974	0.3540	9.1566	0.0843
14	0.8434	0.2752	9.4318	0.0568
15	0.5682	0.2032	9.6350	0.0365
16	0.3650	0.1420	9.7770	0.0223
17	0.2230	0.0935	9.8705	0.0129
18	0.1295	0.0581	9.9286	0.0071
19	0.0714	0.0340	9.9625	0.0037
20	0.0375	0.0188	9.9813	0.0019
21	0.0187	0.0098	9.9911	0.0009

6.2.2 Poisson Computer Program. Relatively simple programs can be written in BASIC computer language to solve the Poisson equation shown below:

$$P = \frac{A^N/N!}{1 + A + A^2/2! + A^3/3! + \cdots} = \frac{A^N/N!}{e^A} = \frac{A^N}{N!}e^{-A}$$

Following is a sample program which can be used for traffic loads up to 86 erlangs. (Loads greater than 86 erlangs may cause register overflow in some computers.) It is written using GWBASIC, Version 3.11.

```
 10  PRINT “POISSON (86 ERLANG MAXIMUM)”
 20  PRINT “P = PROBABILITY OF N TRUNKS BUSY”
 30  PRINT “P1 = PROBABILITY OF BLOCKING”
 40  INPUT “OFFERED TRAFFIC IN ERLANGS = ”, A: LET E = EXP(A)
 50  LET N = 0: LET T = 1: LET T1 = 1: LET T2 = 0: LET Z = 15
 60  PRINT “TRUNKS PR (N TRUNKS BUSY) PR (BLOCKING)”
 70  PRINT “---------------------------------------”
 80  LET P1 = 1 - T2: LET T2 = T1/E: LET P = T/E
 90  IF P<.00001 THEN LET Z = N + 14: GOTO 170
100  PRINT USING “###     ”;N,
110  PRINT USING “        #.####        ”;P,P1
120  IF N<Z THEN 180
130  STOP: LET Z = Z + 15
140  PRINT “OFFERED TRAFFIC IN ERLANGS IS ”;A
150  PRINT “TRUNKS PR (N TRUNKS BUSY) PR (BLOCKING)”
160  PRINT “---------------------------------------”
170  IF N<Z THEN 120: LET Z = Z + 15
180  LET N = N + 1: LET T = T * A/N: LET T1 = T1 + T
190  IF P1>.0001 THEN 80
200  END
```

This program calculates both (a) the probability that exactly “N” trunks are busy and (b) the probability of blocking (the probability that “N” or more service requests are received). The results are presented on a trunk-by-trunk basis until the probability of blocking reaches some minimum value [P1 = Pr(Blocking) = 0.0001 in this version]. Both the minimum and maximum probabilities can be adjusted by changing the values of P (Instruction 90) and P1 (Instruction 190).

The variable “Z” is used to stop the calculation routine every “Z” lines so that the data will fill one screen at a time. Calculation will continue when the instruction “CONT” is entered (or by pressing the “F5” key in some computers).

Following is a sample program calculation for an offered traffic load of 10 erlangs:

```
POISSON (86 ERLANG MAXIMUM)
P = PROBABILITY OF N TRUNKS BUSY
P1 = PROBABILITY OF BLOCKING
OFFERED TRAFFIC IN ERLANGS = 10
TRUNKS     PR (N TRUNKS BUSY)     PR (BLOCKING)
---------------------------------------------------------------
 0              0.0000              1.0000
 1              0.0005              1.0000
 2              0.0023              0.9995
 3              0.0076              0.9972
 4              0.0189              0.9897
 5              0.0378              0.9707
 6              0.0631              0.9329
 7              0.0901              0.8699
 8              0.1126              0.7798
 9              0.1251              0.6672
10              0.1251              0.5421
11              0.1137              0.4170
12              0.0948              0.3032
13              0.0729              0.2084
14              0.0521              0.1355
15              0.0347              0.0835
16              0.0217              0.0487
17              0.0128              0.0270
18              0.0071              0.0143
19              0.0037              0.0072
20              0.0019              0.0035
21              0.0009              0.0016
22              0.0004              0.0007
23              0.0002              0.0003
24              0.0001              0.0001
25              0.0000              0.0000
```

6.2.3 Erlang C Computer Program. Relatively simple programs can be written in BASIC computer language to solve the Erlang C equation shown below:

$$P = \frac{(A^N/N!)(N/(N-A))}{1 + A + A^2/2! + A^3/3! + \cdots + (A^N/N!)(N/(N-A))}$$

Following is a sample program. It is written using GWBASIC, Version 3.11.

```
10 PRINT “ERLANG C CALCULATES DELAY”
20 INPUT “OFFERED TRAFFIC (ERLANGS) = ”,A
30 PRINT “TRUNKS    P    D1    D2    Q1    Q2”;
```

```
40 PRINT "  P8   P4    P2     P1    PP"
50 PRINT
"------------------------------------------------------------------------"
60 LET N = 1: LET T = 1: LET T1 = 1
70 IF T1 > 1E + 21 THEN 200
80 IF N<=A THEN 190
90 IF N = A + 1 THEN LET Z = A + 15
100 LET T2 = T * (A/N) * (N/(N - A)): LET P = T2/(T1 + T2)
110 LET D2 = 1/(N - A): LET D1 = P * D2: LET Q2 = A * D2: LET Q1 = P * Q2
120 LET PO = P/EXP(2/D2)
130 PRINT USING "###      ";N,
140 PRINT USING ".####    ";P,
150 PRINT USING "###.##  ";D1, D2, Q1, Q2,
160 PRINT USING ".##       ";P0
170 IF N>Z THEN 210
180 IF P<.02 THEN 250
190 LET T = T * A/N: LET T1 = T1 + T: LET N = N + 1: GOTO 70
200 LET T = T * 1E - 37: LET T1 = T1 * 1E - 37: GOTO 80
210 STOP: LET Z = Z + 15
220 PRINT "OFFERED TRAFFIC (ERLANGS) = ",A
230 PRINT "TRUNKS    P    D1    D2     Q1    Q2     PP"
240 PRINT "------------------------------------": GOTO 180
250 END
```

For a given traffic load, this program shows the probability of delay for a group with "N" trunks, with call delays expressed in units of average holding time:

P	probability that a call will be delayed
D1	average delay on all calls, including those not delayed
D2	average delay on calls that are delayed
Q1	average calls in queue
Q2	average calls in queue when all servers are busy
PP	probability of delay exceeding two holding time intervals

The program calculates delay and queue characteristics by using the following relationships derived from the Erlang C formula:

The average delay on all calls, D1, including those not delayed, is given by

$$\mathrm{D1} = \mathrm{P}(>0)(\mathrm{h}/(\mathrm{N} - \mathrm{A})), \qquad \text{where h is the average holding time}$$

The average delay, D2, on calls delayed is given by

$$\mathrm{D2} = \mathrm{h}/(\mathrm{N} - \mathrm{A}), \qquad \text{where h is the average holding time}$$

The average calls in queue, Q2, when all servers are busy is given by

$$\mathrm{Q2} = \mathrm{Ah}/(\mathrm{N} - \mathrm{A}) = \mathrm{AD2}$$

The average calls in queue, Q1, is given by

$$\mathrm{Q1} = \mathrm{P}(>0)(\mathrm{Ah}/(\mathrm{N}-\mathrm{A})) = \mathrm{P}(>0)\mathrm{Q2}$$

The probability of delay exceeding some multiple of the average holding time, PP (PP $> 2 * \mathrm{HT}$ is used in the program), is given by

$$\mathrm{PP} = \mathrm{P}(>0)/\mathrm{e}^{(\mathrm{XHT}/\mathrm{D2})}, \qquad \text{where X is the number of HT intervals}$$

This assumes that the trunks are used in sequential order. The results are presented on a trunk-by-trunk basis beginning with "A + 1" trunks to ensure that there are sufficient trunks available to carry the busy hour traffic after some delay has occurred. This process continues until the probability of delay is small enough that delays are relatively short [Pr(Delay) = .02 in this version]. The minimum delay probability can be adjusted as desired.

The variable "Z" is used to stop the calculation routine every "Z" lines so that the data will fill one screen at a time. Calculation will continue when the instruction "CONT" is entered (or by pressing the "F5" key in some computers).

In order to avoid overflow in the computer registers when very large numbers are encountered, the least significant digits are dropped for large traffic loads (see Instruction 70 and Instruction 200). The variables affected by this adjustment (T and T1) are used only to determine the probability ratio, and the error introduced is negligible. As the traffic load A becomes very large, the error increases.

Following is a sample program calculation for an offered traffic load of 10 erlangs:

```
ERLANG C CALCULATES DELAY
OFFERED TRAFFIC (ERLANGS) = 10
TRUNKS      P       D1      D2       Q1       Q2       PP
------------------------------------------------------------------------
11        0.6821   0.68    1.00     6.82    10.00    0.09
12        0.4494   0.22    0.50     2.25     5.00    0.01
13        0.2853   0.10    0.33     0.95     3.33    0.00
14        0.1741   0.04    0.25     0.44     2.50    0.00
15        0.1020   0.02    0.20     0.20     2.00    0.00
16        0.0573   0.01    0.17     0.10     1.67    0.00
17        0.0309   0.00    0.14     0.04     1.43    0.00
18        0.0159   0.00    0.13     0.02     1.25    0.00
```

Source: The information in Section 6.2 was graciously provided by John Lawlor Associates, Sharon, MA.

7 WAITING SYSTEMS (QUEUEING)

A short discussion follows regarding traffic in queueing systems. Queueing or waiting systems, when dealing with traffic, are based on the third assumption, namely, lost calls delayed (LCD). Of course, a queue in this case is a pool of callers waiting to be served by a switch. The term *serving time* is the time a call takes to be served from the moment of arrival in the queue to the moment of being served by the switch. For traffic calculations in most telecommunication queueing systems, the mathematics is based on the assumption that call arrivals are random and Poissonian. The traffic engineer is given the parameters of offered traffic, the size of the queue, and a specified grade of service and will determine the number of serving circuits or trunks required.

The method by which a waiting call is selected to be served from the pool of waiting calls is called *queue discipline*. The most common discipline is the first-come, first-served discipline, where the call waiting longest in the queue is served first. This can turn out to be costly because of the equipment required to keep order in the queue. Another type is random selection, where the time a call has waited is disregarded and those waiting are selected in random order. There is also the last-come, first-served discipline and bulk service discipline, where batches of waiting calls are admitted, and there are also priority service disciplines, which can be preemptive and nonpreemptive. In queueing systems the grade of service may be defined as the probability of delay. This is expressed as $P(t)$, the probability that a call is not being immediately served and has to wait a period of time greater than t. The average delay on all calls is another parameter that can be used to express grade of service, and the length of queue is another.

The probability of delay, the most common index of grade of service for waiting systems when dealing with full availability and a Poissonian call arrival process, is calculated by using the Erlang C formula, which assumes an infinitely long queue length. Syski [3] provides a good guide to Erlang C and other, more general waiting systems. (Also consult Refs. 13–15.)

7.1 Server-Pool Traffic

Server pools are groups of traffic resources, such as signaling registers and operator positions, that are used on a shared basis. Service requests that cannot be satisfied immediately are placed in a queue and served on a first-in, first-out (FIFO) basis. Server-pool traffic is directly related to offered traffic, server-holding time, and call-attempt factor and is inversely related to call-holding time. This is expressed in equation 1.14.

$$A_S = \frac{A_T \cdot T_S \cdot C}{T_C} \tag{1.14}$$

where A_S = server-pool traffic in erlangs
A_T = total traffic served in erlangs
T_S = mean server-holding time in hours
T_C = mean call-holding time in hours
C = call-attempt factor (dimensionless)

Total traffic served refers to the total offered traffic that requires the services of the specific server pool for some portion of the call. For example, a dual-tone multifrequency (DTMF) receiver pool is dimensioned to serve only the DTMF tone-dialing portion of total switch traffic generated by DTMF signaling sources.

Table 1.1 gives representative server-holding times for typical signaling registers as a function of the number of digits received or sent.

The mean server-holding time is the arithmetic average of all server-holding times for the specific server pool. Equation 1.15 can be used to calculate mean server-holding time for calls with different holding-time characteristics.

$$T_S = a \cdot T_1 + b \cdot T_2 + \cdots + k \cdot T_n \tag{1.15}$$

where T_S = mean server-holding time in hours
$T_1, T_2, \ldots, T_n$ = individual server-holding times in hours
$a, b, \ldots, k$ = fractions of total traffic served ($a + b + \cdots + k = 1$)

Consider the following example. Determine the mean DTMF receiver-holding time for a switch where subscribers dial local calls using a 7-digit number and long-distance calls using an 11-digit number. Assume that 70% of the calls are local calls, the remainder are long-distance calls, and the typical signaling register holding times found in Table 1.1 are applicable.

$$T_s = (0.7)(8.1\text{ s}) + (0.3)(12.0\text{ s}) = 9.27\text{ s}$$

Call-attempt factors are dimensionless numbers that adjust offered traffic intensity to compensate for call attempts that do not result in completed calls. Therefore, call-attempt factors are inversely proportional to the fraction of completed calls as defined in equation 1.16.

$$C = 1/k \tag{1.16}$$

where C = call-attempt factor (dimensionless)
k = fraction of calls completed (decimal fraction)

Here is another example. Table 1.2 gives representative subscriber call-attempt dispositions based on empirical data taken from the large North American PSTN

TABLE 1.1 Typical Signaling Register Holding Times in Seconds

	Number of Digits Received or Sent				
Signaling Register	1	4	7	10	11
Local dial-pulse (DP) receiver	3.7	8.3	12.8	17.6	19.1
Local DTMF receiver	2.3	5.2	8.1	11.0	12.0
Incoming MF receiver	1.0	1.4	1.8	2.2	2.3
Outgoing MF sender	1.5	1.9	2.3	2.8	3.0

TABLE 1.2 Typical Call-Attempt Dispositions

Call-Attempt Disposition	Percentage
Call was completed	70.7
Called subscriber did not answer	12.7
Called subscriber line was busy	10.1
Call abandoned without system response	2.6
Equipment blockage or failure	1.9
Customer dialing error	1.6
Called directory number changed or disconnected	0.4

data base. Determine the call-attempt factors for these data, where 70.7% of the calls were completed ($k = 0.707$).

$$C = \frac{1}{k} = \frac{1}{0.707} = 1.414$$

Source: Section 7.1 is based on Section 1.2.2 of *Traffic System Design Handbook* by James R. Boucher, IEEE Press, 1992 [26].

8 DIMENSIONING AND EFFICIENCY

By definition, if we were to dimension a route or estimate the required number of servicing channels, where the number of trunks (or servicing channels) just equaled the erlang load, we would attain 100% efficiency. All trunks would be busy with calls all the time or at least for the entire BH. This would not even allow several moments for a trunk to be idle while the switch decided the next call to service. In practice, if we engineered our trunks, trunk routes, or switches this way, there would be many unhappy subscribers.

On the other hand, we do, indeed, want to size our routes (and switches) to have a high efficiency and still keep our customers relatively happy. The goal of our previous exercises in traffic engineering was just that. The grade of service is one measure of subscriber satisfaction. As an example, let us assume that between cities X and Y there are 100 trunks on the interconnecting telephone route. The tariffs, from which the telephone company derives revenue, are a function of the erlangs of carried traffic. Suppose we allow a dollar per erlang-hour. The very upper limit of service on the route is 100 erlangs. If the route carried 100 erlangs of traffic per day, the maximum return on investment would be $2400 a day for that trunk route and the portion of the switches and local plant involved with these calls. As we well know, many of the telephone company's subscribers would be unhappy because they would have to wait excessively to get calls through from X to Y. How, then, do we optimize a trunk route (or serving circuits) and keep the customers as happy as possible?

In our previous discussions, an excellent grade of service was 0.001. We relate grade of service to subscriber satisfaction. Turning to Ref. 20, Table 1-2, such a grade of service with 100 circuits would support 75.24 erlangs during the BH. With 75.24 erlangs loading, the route would earn \$75.24 during that one-hour period and something far less that \$2400 per day. If the grade of service was reduced to 0.01, 100 trunks would bring in \$84.06 for the busy hour. Note the improvement in revenue at the cost of reducing grade of service. Another approach to save money is to hold the erlang load constant and decrease the number of trunks and switch facilities accordingly as the grade of service is reduced. For instance, 70 erlangs of traffic at $p = 0.001$ requires 96 trunks and at $p = 0.01$, only 86 trunks.

8.1 Alternative Routing

One method of improving efficiency is to use alternative routing (called *alternate routing* in North America). Suppose that we have three serving areas, X, Y, and Z, served by three switches, X, Y, and Z as illustrated in Figure 1.8.

Let the grade of service be 0.005 (1 in 200 in Table 1-2, Ref. 20). We found that it would require 67 trunks to carry 50 erlangs of traffic during the BH to meet that grade of service between X and Y. Suppose that we reduced the number of trunks between X and Y, still keeping the BH traffic intensity at 50 erlangs. We would thereby increase the efficiency on the $X-Y$ route at the cost of reducing the grade of service. With a modification of the switch at X, we could route the traffic bound for Y that met congestion on the $X-Y$ route via Z. Then Z would route this traffic on the $Z-Y$ link. Essentially, this is alternative routing in its simplest form. Congestion probably would only occur during very short peaky periods in the BH, and chances are that these peaks would not occur simultaneously with peaks in traffic intensity on the $Z-Y$ route. Furthermore, the added load on the $X-Z-Y$ route would be very small. Some idea of traffic peakiness that would overflow onto the secondary route $(X + Z + Y)$ is shown in Figure 1.9.

One of the most accepted methods of dimensioning switches and trunks using alternative routing is the equivalent random group (ERG) method developed by Wilkinson [11]. The Wilkinson method uses the mean M and the variance V.

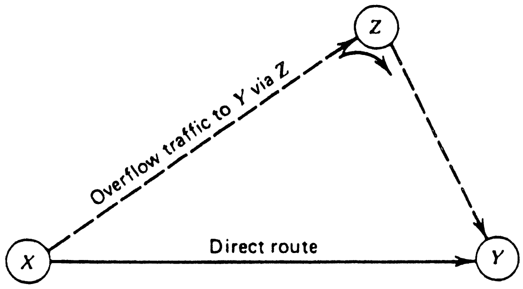


Figure 1.8. Simplified diagram of the alternative routing concept (solid line represents direct route, dashed line represents alternative route carrying the overflow from X to Y).

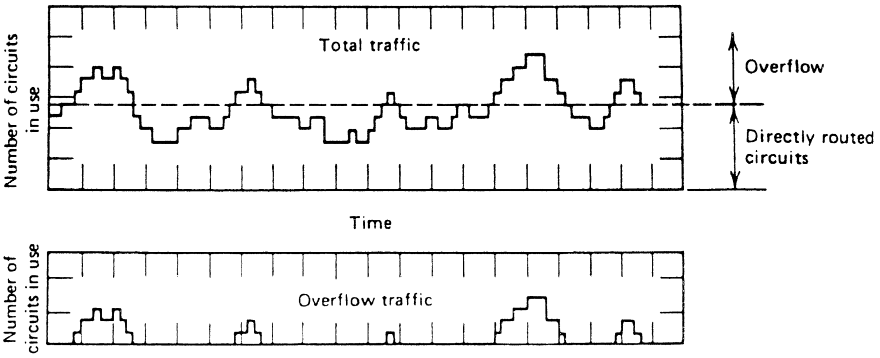


Figure 1.9. Traffic peakiness, the peaks representing overflow onto alternative routes.

Here the *overflow traffic* is the "lost" traffic in the Erlang B calculations, which were discussed earlier. Let M be the mean value of that overflow and A be the random traffic offered to a group of n circuits (trunks). Then

$$V = M\left(1 - M + \frac{A}{1 + n + M - A}\right) \tag{1.17}$$

When the overflow traffic from several sources is combined and offered to a single second (or third, fourth, etc.) choice of a group of circuits, both the mean and the variance of the combined traffic are the arithmetical sums of the means and variances of the contributors.

The basic problem in alternative routing is to optimize circuit group efficiency (e.g., to dimension a route with an optimum number of trunks). Thus we are to find what circuit quantities result in minimum cost for a given grade of service, or to find the optimum number of circuits (trunks) to assign to a direct route allowing the remainder to overflow on alternative choices. There are two approaches to the optimization. The first method is to solve the problem by successive approximations, and this lends itself well to the application of the computer [12]. Then there are the manual approaches, two of which are suggested in CCITT Rec. E.525 [25]. Alternate (alternative) routing is further discussed in Chapter 6.

8.2 Efficiency versus Circuit Group Size

In the present context a *circuit group* refers to a group of circuits performing a specific function. For instance, all the trunks (circuits) routed from X to Y in Figure 1.8 make up a circuit group, irrespective of size. This group should not be confused with the "group" used in transmission-engineering carrier multiplex systems.

If we assume full loading, it can be stated that efficiency improves with circuit group size. From Table 1-2 of Ref. 20, given $p = 0.001$, 5 erlangs of traffic

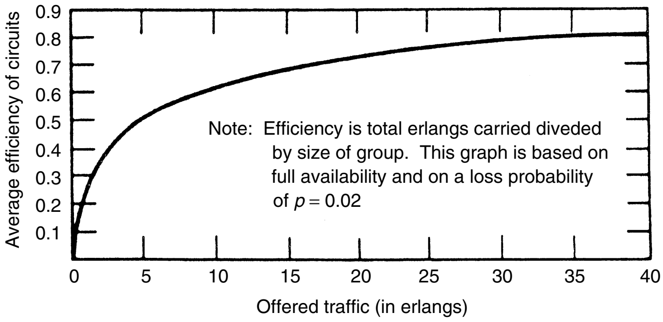


Figure 1.10. Group efficiency increases with size.

requires a group with 11 trunks, more than a 2 : 1 ratio of trunks to erlangs, and 20 erlangs requires 30 trunks, a 3 : 2 ratio. Note how the efficiency has improved. One hundred twenty trunks will carry 100 erlangs, or 6 trunks for every 5 erlangs for a group of this size. Figure 1.10 shows how efficiency improves with group size.

9 BASES OF NETWORK CONFIGURATIONS

In this section we discuss basic network configurations that may apply anywhere in the telecommunication community. Networks more applicable to the local area are covered in Chapter 2, and those for the long-distance plant are discussed in Chapter 6.

9.1 Introductory Concepts

A network in telecommunications may be defined as a method of connecting exchanges so that any one subscriber in the network can communicate with any other subscriber. For this introductory discussion, let us assume that subscribers access the network by a nearby local exchange. Thus the problem is essentially how to connect exchanges efficiently. There are three basic methods of connection in conventional telephony: (1) mesh, (2) star, and (3) double and higher-order star (see Section 2 of this chapter). The mesh connection is one in which each and every exchange is connected by trunks (or junctions) to each and every other exchange as shown in Figure 1.11A. A star connection utilizes an intervening exchange, called a *tandem exchange*, such that each and every exchange is interconnected via a *single* tandem exchange. An example of a star connection is shown in Figure 1.11B. A double-star configuration is one where sets of pure star subnetworks are connected via higher-order tandem exchanges, as shown in Figure 1.11C. This trend can be carried still further, as we see later on, when hierarchical networks are discussed.

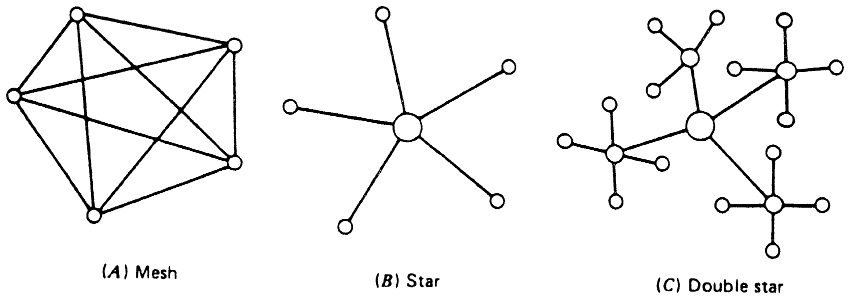
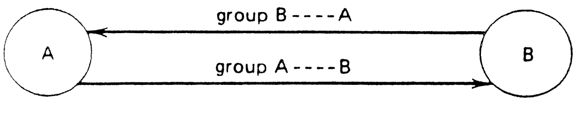


Figure 1.11. Examples of star, double-star, and mesh configurations.

As a general rule we can say that mesh connections are used when there are comparatively high traffic levels between exchanges, such as in metropolitan networks. On the other hand, a star network may be applied when traffic levels are comparatively low.

Another factor that leads to star and multiple-star network configurations is network complexity in the trunking outlets (and inlets) of a switch in a full mesh. For instance, an area with 20 exchanges would require 380 traffic groups (or links), and an area with 100 exchanges would require 9900 traffic groups. This assumes what are called *one-way groups*. A one-way group is best defined considering the connection between two exchanges, *A* and *B*. Traffic originating at *A* bound for *B* is carried in one group and the traffic originating at *B* bound for *A* is carried in another group, as shown in the following diagram:



One-way and both-way groups are further discussed in Section 11.

Thus, in practice, most networks are compromises between mesh and star configurations. For instance, outlying suburban exchanges may be connected to a nearby major exchange in the central metropolitan area. This exchange may serve nearby subscribers and be connected in mesh to other large exchanges in the city proper. Another example is the city's long-distance exchange, which is a tandem exchange looking into the national long-distance network, whereas the major exchanges in the city are connected to it in mesh. An example of a real-life compromise among mesh, star, and multiple-star configurations is shown in Figure 1.12.

9.2 Higher-Order Star Network

Figure 1.13 illustrates a higher-order star network. It is simply several star networks of Figure 1.11B stacked on top of each other. Another high-order star

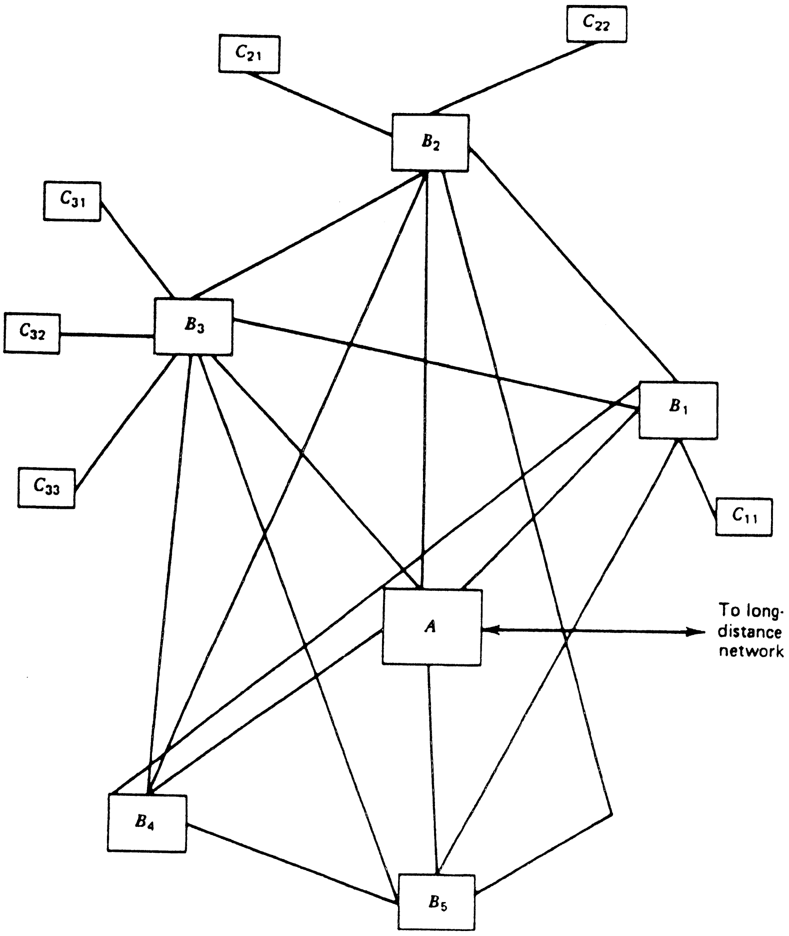


Figure 1.12. A typical telephone network serving a small city as an example of a compromise between mesh and star configuration. *A* is the highest level in this simple hierarchy. *A* might house the "point of presence" (POP) in the U.S. network. *B* is a local exchange. *C* may be a satellite exchange or a concentrator. Consult Ref. 24.

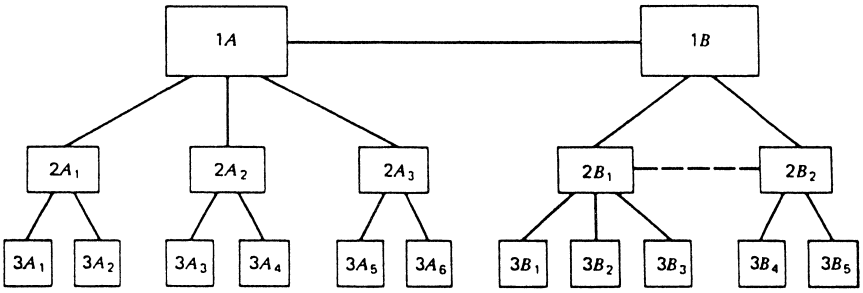


Figure 1.13. Higher-order star network.

network is shown in Figure 1.11C. In these types of networks the higher layers are given more importance than the lower layers. The incisive reader will say that we describe a hierarchical network. The reader is correct, but we wish to reserve our detailed discussion of hierarchical networks for Chapter 6, Sections 6 and 7.

We illustrate the order of importance of the several levels in a high-order star network in Figure 1.13. There are three levels or ranks of exchanges in the figure. The smallest blocks in the diagram are the lowest-ranked exchanges, which have been marked with a "3" to indicate the third level or rank. Note that there are restrictions or rules of traffic flow. As the figure is drawn, traffic from $3A_1$ to $3A_2$ would have to flow through exchange $2A_1$. Likewise, traffic from exchange $2A_2$ to $2A_3$ would have to flow through exchange 1A. Carrying the concept one step further, traffic from any A exchange to any B exchange would necessarily have to be routed through exchange 1A.

The next consideration is the high-usage (HU) route. For instance, if we found that there were high traffic intensities (e.g., >20 erlangs) between $2B_1$ and $2B_2$, trunks and switch gear might well be saved by establishing a HU route between the two (shown by a dashed line in Figure 1.13). Thus we might call the high-usage route a *highly traveled shortcut*. Of course, HU routes could be established between any pair of exchanges in the network if traffic intensities and distances involved proved this strategy economical. When HU routes are established, traffic between the exchanges involved will first be offered to the HU route, and overflow would take place through a last choice route or, as shown in Figure 1.13, up to the next level and down. If routing is through the highest level of higher-order star network, we call this route the final route, a hierarchical network term. (See Chapter 6, Sections 6 and 7.)

10 VARIATIONS IN TRAFFIC FLOW

In networks covering large geographic expanses and even in cases of certain local networks, there may be a variation in time of day of the BH or in the direction of traffic flow. In the United States, business traffic peaks during several hours before and after the noon lunch period on weekdays, and social calls peak in early evening. Traffic flow tends to be from suburban living areas to urban centers in the morning, and the reverse in the evening.

In national networks covering several time zones where the differences in local time may be appreciable, long-distance traffic tends to be concentrated in a few hours common to BH peaks at both ends. In such cases it is possible to direct traffic so that peaks of traffic in one area fall into valleys of traffic in another (noncoincident busy hour). The network design can be made more economical if configured to take advantage of these phenomena, particularly in the design and configuration of direct routes versus overflow.

11 ONE-WAY AND BOTH-WAY (TWO-WAY) CIRCUITS

We defined one-way circuits in Section 9.1. Here traffic from A to B is assigned to one group of circuits, and traffic from B to A is assigned to another separate group. In both-way (or two-way) operation a circuit group may be engineered to carry traffic in both directions. The individual circuits in the group may be used in either direction, depending on which exchange seizes the circuit first.

In engineering networks it is most economical to have a combination of one-way and both-way circuits on longer routes. Signaling and control arrangements on both-way circuits are substantially more expensive. However, when dimensioning a system for a given traffic intensity, fewer circuits are needed in both-way operation, with notable savings on low-intensity routes (i.e., below about 10 erlangs in each direction). For long circuits, both-way operation has obvious advantages when dealing with a noncoincident BH. During overload conditions, both-way operation is also advantageous because the direction of traffic flow in these conditions is usually unequal.

The major detriment to two-way operation, besides its increased signaling cost, is the possibility of double seizure. This occurs when both ends seize a circuit at the same time. There is a period of time when double seizure can occur in a two-way circuit; this extends from the moment the circuit is seized to send a call and the moment when it becomes blocked at the other end. Signaling arrangements can help to circumvent this problem. Likewise, switching arrangements can be made such that double seizure can occur only on the last free circuit of a group. This can be done by arranging in turn the sequence of scanning circuits so that the sequence on one end of a two-way circuit is reversed from that of the other end. Of course, great care must be taken on circuits having long propagation times, such as satellite and long undersea cable circuits. By extending the time between initial seizure and blockage at the other end, these circuits are the most susceptible, just because a blocking signal takes that much longer to reach the other end.

12 QUALITY OF SERVICE

Quality of service appears at the outset to be an intangible concept. However, it is very tangible for a telephone subscriber unhappy with his or her service. The concept of service quality must be mentioned early in any all-encompassing text on telecommunications systems. System engineers should never once lose sight of the concept, no matter what segment of the system they may be responsible for. Quality of service also means *how happy* the telephone company (or other common carrier) is keeping the customer. For instance, we might find that about half the time a customer dials, the call goes awry or the caller cannot get a dial tone or cannot hear what is being said by the party at the other end. All these have an impact on quality of service. So we begin to find that quality of service is an important factor in many areas of the telecommunications business and means different things to different people. In the old days of telegraphy, a rough measure of how well the system was working was the number of service messages received at a switching center. In modern telephony we now talk about service observing (see Chapter 3, Section 17).

The transmission engineer calls quality of service "customer satisfaction," which is commonly measured by how well the customer can hear the calling party. It is called *loudness rating* (see Chapter 2, Section 2.2.1) and is measured in decibels. In our discussion of traffic, lost calls or blockage certainly constitute another measure of service quality. If this is measured in a decimal quantity, one target figure for grade of service would be $p = 0.01$. Other items listed under service quality are:

- Delay before receiving dial tone ("dial-tone delay").
- Postdial(ing) delay (time from completion of dialing a number to first ring of telephone called).
- Availability of service tones (busy tone, telephone out of order, ATB, etc.).
- Correctness of billing.
- Reasonable cost to customer of service.
- Responsiveness to servicing requests.
- Responsiveness and courtesy of operators.
- Time to installation of new telephone, and, by some, the additional services offered by the telephone company [22].

One way or another, each item, depending on service quality goal, will have an impact on the design of the system.

Furthermore, each item on the list can be quantified, usually statistically, such as loudness rating, or in time, such as time taken to install a telephone. In some countries it can be measured in years. Regarding service quality, good reading can be found in ITU-T Recs. E.430, E.432, I.350, and X.140.

REVIEW QUESTIONS

1. Give the standard telephone battery voltage with respect to ground.
2. What is *on hook* and *off hook*? When a subscriber subset (the telephone) goes "off hook," what occurs at the serving switch? List two items.
3. Suppose that the sidetone level of a telephone is increased. What is the natural reaction of the subscriber?
4. A subscriber pair, with a fixed battery voltage, is extended. As we extend the loop further, two limiting performance factors come into play. Name them.

5. Define a mesh connection. Draw a star arrangement.

6. In the context of the argument presented in the chapter, what is the principal purpose of a local switch?

7. What are the two basic parameters that define “traffic”?

8. Distinguish offered traffic from carried traffic.

9. Give one valid definition of the *busy hour*.

10. On a particular traffic relation the calling rate is 461 and the average call duration is 1.5 min during the BH. What is the traffic intensity in CCS, in erlangs?

11. Define *grade of service*.

12. A particular exchange has been dimensioned to handle 1000 calls during the busy hour. On a certain day during the BH 1100 calls are offered. What is the resulting grade of service?

13. Distinguish a full availability switch from a limited availability switch.

14. In traffic theory there are three ways lost calls are handled. What are they?

15. Call arrivals at a large switch can be characterized by what type of mathematical distribution? Such arrivals are________in nature.

16. Based on the Erlang B formula and given a BH requirement for a grade of service of 0.005 and a BH traffic intensity of 25 erlangs on a certain traffic relation, how many trunks are required?

17. Carry out the same exercise as in question 16 but use the Poisson tables to determine the number of trunks required.

18. Give at least two queueing disciplines.

19. As the grade of service is improved, what is the effect on trunk efficiency?

20. What is the basic purpose of alternative routing? What does it improve?

21. How does circuit group size (number of trunks) affect efficiency for a fixed grade of service?

22. Give the three basic methods of connecting exchanges. (These are the three basic network types.)

23. At what erlang value on a certain traffic relation does it pay to use tandem routing? Is this a maximum or a minimum value?

24. Differentiate between one-way and both-way circuits.

25. What is the drawback of one-way circuits? of both-way circuits?

26. Hierarchical networks are used universally in national and international telephone networks. Differentiate between high-usage (HU) connectivities and final route.

27. Distinguish a tandem exchange from a transit exchange.

28. Name at least five items that can be listed under *quality of service* (QoS).

REFERENCES

1. International Telephone and Telegraph Corporation, *Reference Data for Radio Engineers*, 5th ed., Howard W. Sams, Indianapolis, 1968.
2. R. R. Mina, "The Theory and Reality of Teletraffic Engineering," *Telephony*, a series of articles (April 1971).
3. R. Syski, *Introduction to Congestion Theory in Telephone Systems*, Oliver and Boyd, Edinburgh, 1960.
4. G. Dietrich et al., *Teletraffic Engineering Manual*, Standard Electric Lorenz, Stuttgart, Germany, 1971.
5. E. Brockmeyer et al., "The Life and Works of A. K. Erlang," *Acta Polytechnica Scandinavia*, The Danish Academy of Technical Sciences, Copenhagen, 1960.
6. *A Course in Telephone Traffic Engineering*, Australian Post Office, Planning Branch, 1967.
7. Arne Jensen, *Moe's Principle*, The Copenhagen Telephone Company, Copenhagen, Denmark, 1950.
8. *Networks*, Laboratorios ITT de Standard Eléctrica SA, Madrid, 1973 (limited circulation).
9. *Local Telephone Networks*, The International Telecommunications Union, Geneva, 1968.
10. *Electrical Communication System Engineering Traffic*, U.S. Department of the Army, TM-11-486-2, August 1956.
11. R. I. Wilkinson, "Theories for Toll Traffic Engineering in the USA," *BSTJ*, **35** (March 1956).
12. *Optimization of Telephone Trunking Networks with Alternate Routing*, ITT Laboratories of Standard Eléctrica (Spain), Madrid, 1974 (limited circulation).
13. J. Riordan, *Stochastic Service Systems*, John Wiley & Sons, New York, 1962.
14. L. Kleinrock, *Queueing Systems*, Vols. 1 and 2, John Wiley & Sons, New York, 1975.
15. T. L. Saaty, *Elements of Queueing Theory with Applications*, McGraw-Hill, New York, 1961.
16. J. E. Flood, *Telecommunication Networks*, IEE Telecommunications Series 1, Peter Peregrinus, London, 1975.
17. "Telcordia Notes on the Networks," Telecordia Special Report SR-2275, Issue 4, Telecordia, Piscataway, NJ, October 2000.
18. *National Telephone Networks for the Automatic Service*, International Telecommunication Union—CCITT, Geneva, 1964.
19. D. Bear, *Principles of Telecommunication Traffic Engineering*, IEE Telecommunications Series 2, Peter Peregrinus, London, 1976.
20. R. L. Freeman, *Reference Manual for Telecommunication Engineering*, 3rd ed., John Wiley & Sons, New York, 2002.
21. *Traffic Routing*, CCITT Rec. E.170, ITU Geneva, October 1992.

22. "Telecommunications Quality," *IEEE Communications Magazine* (entire issue), IEEE, New York, 1988.
23. *Engineering and Operations in the Bell System*, 2nd ed., AT&T Bell Laboratories, Murray Hill, NJ, 1983.
24. *BOC Notes on the LEC Networks—1994*, Issue 2, SR-TSV-002275, Bellcore, Piscataway, NJ, April 1994.
25. Private Communications—John Lawlor/John Lawlor Associates, Sharon, MA, December 1994.
26. J. R. Boucher, *Traffic System Design Handbook*, IEEE Press, New York, 1992.
27. *Designing Networks to Control Grade of Service*, CCITT Rec. E.525, ITU, Geneva, 1992.