Chapter 11
Telephone System Overview

The world's telephone system has gone through monumental changes in the last thirty or so years. Under the pressure of competing companies, and new technology, telephone companies around the world have overhauled the system to the point where there is little similarity to what existed fifty or more years ago.

But because the financial investment in the telephone network is so huge, these changes have happened slowly and piecemeal. Wherever possible, each new technique or piece of equipment has been designed to be compatible with older equipment, so that old and new could coexist during the transition period. In fact, there are still places in the world where very old equipment exists. In this chapter, we provide an overall view of both old and new telephone technology, and describe how the upgrading was done. We will discuss less of the technical details, and more of overall concepts behind the system, including historical and regulatory concepts.

POTS History

POTS is the abbreviation for the *Plain Old Telephone System* — also sometimes called PSTN for *Public Switched Telephone Network*. We can describe the history of the system as follows:

The telephone was simultaneously and independently invented about 1870–76 by Alexander Graham Bell and Elisha Gray. The men were competitors, but Bell patented the invention first and therefore got the credit.

The initial telephones were used for point-to-point communications between two places; the first telephone switchboard was not invented until the early 1880's. It was manual, and required the services of an operator to make the connection. The first automatic switchboard (and dial telephone) was invented in the early 1890's. This allowed the automatic dialing of local calls, but operators were still needed for long distance connections as late as the 1950's.

The period from 1870 through 1913 was a period of great competition between rival telephone companies. In many cases, multiple telephone companies served the same area, and a subscriber to one company could not talk to the subscriber of another. The U.S. Department of Justice eventually stepped into the case, and forced the largest company (AT&T) to interconnect with the others.

By 1950 or so, local telephone service was provided in the U.S. by more than 1000 small telephone companies (some of which only serviced one town), and a group of large telephone companies owned by AT&T (which were called the Bell Companies, such as New York Bell Telephone Company). Long distance service was provided by AT&T itself.

Despite being connected to one another, these companies all had regulations forbidding connection to anyone else. For example, even placing a plastic cover on a telephone directory was considered the "connection of foreign equipment" and forbidden in some parts of the country. Homeowners had to rent a telephone from the local telephone company, and were not allowed to own their own. Telephone answering machines were illegal. As a result, several companies sued AT&T and won. In 1968, it became legal to connect your own telephone or modem to telephone lines, although with restrictions and some difficulty.

The process changed in 1984 when a court decision called the *Modified Final Judgment* forced AT&T to break up. The breakup is called the "divestiture" because it forced AT&T to divest itself of its local telephone companies. This has been further modified with various later court decisions and government actions, so that the current situation is like this:

- AT&T provides long distance service. In addition, there are other long distance companies, such as MCI, which also provide long distance connections. They are called IEC's or IXC's for *Inter-Exchange Carriers*.
- The original Bell Telephone Companies, now called BOC or *Bell Operating Companies*, have been separated (sold off) from AT&T, and grouped together into seven groups called RBOC or *Regional Bell Operating Companies*. Some of these have since merged back together and changed names. For example, the New York Telephone Company and the New England Telephone and Telegraph Company together were the NYNEX RBOC. NYNEX then merged with Bell Atlantic in the middle-Atlantic states, and kept the Bell Atlantic name. Bell Atlantic then merged with General Telephone and Electronics (GTE), an independent, and became Verizon.
- Most of the independent local companies still exist, but some of them have merged together, or in some cases merged with an RBOC. These local companies, as well as the BOCs, are now called LEC or *Local Exchange Carriers*.
- Initially, AT&T and the RBOCs were not allowed to compete with each other — AT&T could not provide local service, and the RBOCs could not provide long distance service. But by the late 1990's, the government allowed them to compete, providing that they allowed other com-

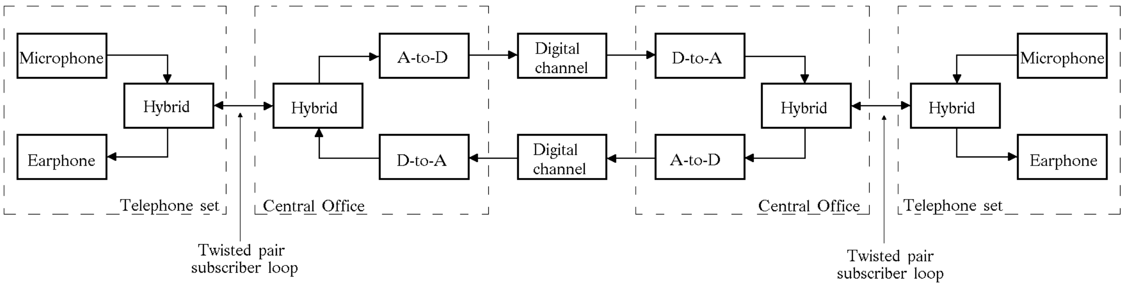

Fig. 11-1. Signal path through the POTS network

panies to compete with them as well. For example, RBOCs were told they could provide long distance service if they allowed other telephone companies — called CLECs or *Competitive Local Exchange Carriers* — to provide local service in their area on an equal basis.

- So we are now back to the case where there are multiple local telephone companies serving the same local area, except that this time they are all interconnected and can call each other. In fact, some of the local telephone companies either lease facilities from an RBOC, or rent space in an RBOC building and place their equipment there in a process called *colocation*. This kind of competition only exists, however, in big cities where there is enough call volume to justify the competition, and where the customers are densely packed. The RBOCs therefore claim that these local telephone companies are skimming off the best and most profitable part of the business, and leaving the rest.
- Whereas the local telephone companies used to own all telephone wiring and telephones, they are now not allowed to own the telephones. Customers may own their own phones, or can rent them from third parties (including AT&T). This now also includes pay telephones — a COCOT is a *Customer Owned Coin Operated Telephone*. Many businesses with pay telephones on their premises actually have them installed by independent COCOT companies, rather than by the local telephone company, because COCOTs pay them for the privilege. Although rates for local calls from COCOTs are regulated by the states, long distance rates are not, and users often find themselves billed for extravagant amounts.

The POTS Network

Fig. 11-1 shows a very simplified diagram of the overall telephone system.

Your home telephone has a microphone (also called a *transmitter* in telephone lingo) and an earphone (called a *receiver*), which convert between sound and electrical signals. The transmitter or microphone picks up your voice and converts it to an electrical signal. The *receiver* or earphone converts the incoming electrical signal back into sound. These connect to the telephone line through a *hybrid*.

More details on the hybrid later; for now, we need only say that the hybrid interfaces the two one-way or *simplex* connections (to the earphone, and from the microphone, altogether comprising a 4-wire connection) to the two-way or *duplex* telephone line which leads to the telephone company's *central office* or *CO* (and which is only a 2-wire connection). The hybrid therefore does a 4-line-to-2-line conversion or (vice versa).

The telephone line, called the *subscriber loop*, is a *twisted pair* — a balanced line consisting of two thin conductors (typically 24-gauge) twisted together to reduce interference — which connects between your telephone and the telephone company's equipment. It is important to remember that this line carries voice simultaneously in both directions.

At the other end of the subscriber loop is the *Central Office* (*CO*), also now called an *end office* or *EO*. The CO contains the switching equipment ("the switch") which switches your calls in response to the number you dial. Here another hybrid splits the two-wire two-way subscriber loop back into two one-directional connections. Years ago, the analog audio signal was transmitted to the destination through an analog channel. Today, the outgoing signal is converted from analog to digital, and stays digital all the way until it gets to the central office at the far end. The box labelled "digital channel" in Fig. 11-1 could be made up of copper cables, optical fibers, microwave relays, or even satellite links.

At the other end of the connection, a digital-to-analog converter converts the digital signal back to analog for transmission through the subscriber loop to the other party's telephone set.

Let us now discuss the system in more detail.

The subscriber loop

The subscriber loop is the two-wire twisted pair of copper wires which goes from your house to the telephone company. At your end, the wire typically ends at a *demarcation jack*, which is the boundary between the inside wire (which you generally own) and the outside wire (which belongs to the phone company.) The demarcation jack provides a convenient point for testing when there are problems — it allows you or the telephone technician to determine whether the problem is in your wiring or theirs.

At the other end, this wire goes to the telephone company's switching equipment.

A few decades ago, your subscriber loop extended all the way back to the central office, though not generally as one continuous piece of wire. The typical subscriber loop has a number of splices along the way (one study says there is an average of 11 splices in a typical subscriber loop), where your single twisted pair connects into a small multi-conductor cable, and then perhaps to larger and larger cables (containing hundreds of other pairs) before it reaches the CO.

But these large cables are expensive, installation is expensive, and space is limited. Today, many subscriber loops go only part of the way toward the CO. Quite often, especially in large office or apartment buildings, or in housing subdivisions, there is multiplexing or switching equipment somewhere between your premises and the CO, which combines your signal, along with those of many others, into a single digital signal on a copper or optical cable going back to the CO.

One of the current buzzwords is "fiber to the curb", which implies that an optical fiber may go directly to your street or neighborhood before it is split up into copper wire subscriber loops to individual homes or offices. Telephone companies are installing fibers partially to lower costs, but also partially because they are looking to the future. Optical fiber has vastly larger bandwidth, and can be used to provide all sorts of new services in the future without requiring that the entire neighborhood be rewired. There is also another reason — copper is expensive, and in some neighborhoods, often stolen. Fiber cable, on the other hand, has no value to thieves and will (hopefully) last longer.

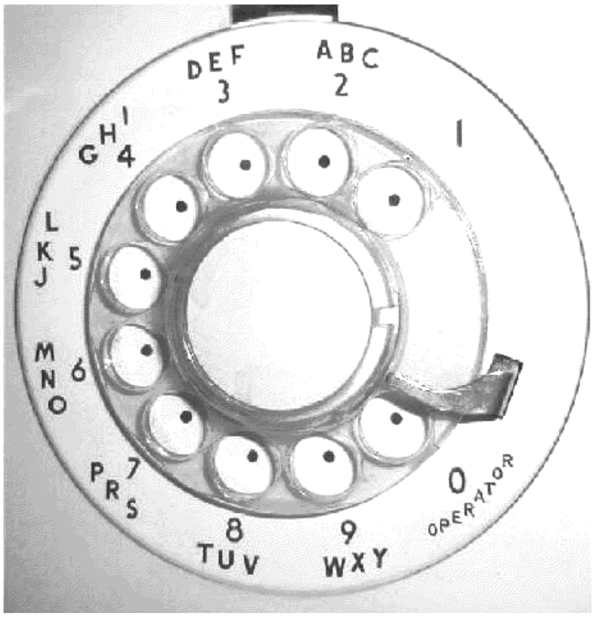

Fig. 11-2. Rotary (pulse) dial

Dialing

Older telephones used a round ("rotary") dial, shown in Fig. 11-2, whereas newer telephones use pushbuttons for dialing. Let us talk about the rotary or pulse dial first.

To dial a number, you place your finger in one of the ten numbered holes, turn the dial clockwise until your finger hits the stop, and then release the dial. Once you release the dial, a spring in the dial turns it back to its starting position. As it does so, a switch contact starts pulsing open-close-open-close... The number of opens depends on the number dialed — when you dial a 1, the switch just opens once; if you dial a 6, it opens six times; if you dial a 0, it opens ten times. Hence the dial is also sometimes called a *pulse* dial.

Each time the dial contact opens, the loop current drops to zero. A circuit in the central office counts the opens to determine what number you dialed. A governor in the dial controls the rotation speed of the dial so that each open-close pulse takes a tenth of a second, to make sure that the central office can properly count the dial pulses. The catch, of course, is that this slows down the dialing. For example, to dial the number 2 may take about $\frac{2}{10}$ of a second (plus whatever time it takes between digits), whereas to dial a 9 would take $\frac{9}{10}$ second (plus the inter-digit time). A complete 7-digit number might easily take five or ten seconds.

Very old central offices used a switching method called *step-by-step*. At that time, each open-close pulse of the dial actually caused a large relay (called a stepping relay or stepper) in the CO to move one step, and

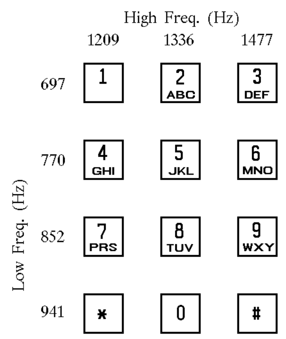

Fig. 11-3. A tone dial and its frequencies

switch from one subscriber (or block of subscribers) to the next. Since the stepper was used for the entire duration of a call, the time it took to dial was unimportant.

In later CO equipment, however, a separate device called a *register* was used to count the dialing pulses. Since dialing only occupies a small part of a typical call, the register was designed to capture the number being dialed, and then release itself to process someone else's call. The longer it took to dial, the longer the register was needed, and therefore, the more registers were needed in the central office. The telephone companies thus needed to speed up the dialing time.

This surfaced when area codes were being defined for long distance dialing. Areas likely to receive many long distance calls, such as New York, Los Angeles, or Chicago, were assigned area codes that were fast to dial — 212, 213, and 312. Locations likely to receive few calls, such as Alaska, Hawaii, or some of the Caribbean islands, were assigned area codes that took longer — such as 907, 808, or 809 (remember that a 0 takes the longest to dial.)

A much more drastic step was to change from pulse dials to tone dials. The Bell System trademark for tone dialing was Touch Tone, while a more generic name is DTMF or *Dual Tone Multi Frequency* dialing. The name Dual Tone Multi Frequency describes the function — there are two (dual) tones, and each one can have several (multi) frequencies.

Fig. 11-3 explains the tones. For example, pressing the number 5 outputs the two frequencies of 770 Hz (for the low tone) and 1336 Hz (for the high tone).

Using tone dialing is much faster than pulse dialing; hence the register in the central office is released much faster than before. Moreover, all digits take equal time to dial, so new area codes can now be assigned without caring whether they are popular or not.

NANP — the North American Numbering Plan

Years ago, a typical central office switch could handle almost 10,000 subscribers. A typical New York City telephone number might be PEnnsylvania 6-4513. *PEnnsylvania 6* described the central office that serviced a particular area, and was called the *exchange*. The first two letters of PEnnsylvania were capitalized to show that they were to be dialed — as Fig. 11-2 or 11-3 show, most of the digits on a dial are assigned three letters; because the letters PE of Pennsylvania were on the digits 7 and 3, PEnnsylvania 6-4513 was actually dialed 736-4513. Mnemonics such as PEnnsylvania 6 or GRamercy 5 were used instead of 736 or 475 because they were judged easier to remember.

Exchange names were often chosen from the name of the area. For example, MUrray Hill 8 was in the Murray Hill area of New York; GRamercy 5 was near Gramercy Park; PEnnsylvania 6 was near the Pennsylvania Railroad station.

The last four digits of the telephone number identified a particular subscriber in that exchange. Except for some numbers which were reserved for other purposes, these four digits could identify up to 10,000 different subscribers. And so the typical exchange had a maximum of 10,000 subscribers.

The problem with exchange names was that it was difficult to assign names to some digit combinations. For example, since the digit 5 corresponds with the letters J, K, and L, it would be hard to find a reasonable name for exchange 552. People might object to names such as KLeptomania 2, or KLooge 2. This was not a problem while the number of exchanges was small, but as the number of telephone lines and exchanges grew, it became important to allow more combinations of exchange numbers, even those with difficult mnemonics. So the telephone companies gradually switched to all-number dialing. Thus PEnnsylvania 6-4513 eventually became just 736-4513.

Although this greatly increased the number of available exchange codes, there were still some limitations. For one, the telephone switching circuits needed some way to tell the difference between an area code and an exchange code. The *NANP* or *North American Numbering Plan* was thus developed with the following rules:

- The first digit of an area code or an exchange could be anything except 1 or 0. The circuitry was designed to ignore an initial 1, in case it was a short pulse due to a loose connection. An initial 0 was not allowed because 0 was used to call the operator.
- To identify an area code, the second digit had to be either 1 or 0. Conversely, exchange codes

were not allowed to have either 0 or 1 in the second position.

- The third digit of either an area code or an exchange could be anything except a 0 or 1. The reason was that three-digit numbers ending in either 1 or 0 were designed for special purposes. For example, you would dial 211 for the long distance operator, 411 for information, 611 for the repair department, and 811 for the telephone company business office. And area code 800 was reserved for "free" calls (called *inWATS* or *inbound Wide Area Telephone Service*) — calls that were paid for by the recipient, rather than the caller.

These rules made it fairly easy for the switching equipment to quickly identify a number as soon as the first few digits were dialed. For example, anything beginning with 212 was destined for New York City's area code 212; anything beginning with 734 had to be an exchange, and so on.

The NANP system worked well for many years. But it still limited the number of area codes and exchanges. Theoretically, three-digit numbers could specify up to 1000 different values. But with the limitations of the NANP, there could be a maximum of just 128 area codes, and just 512 exchanges within any one area code.

By the 1980's, this was just not enough in many locations. For example, all of New York City was initially area code 212, but there were so many telephones in some parts of the city that more than 512 exchanges were needed. So New York was split into two area codes — the island of Manhattan remained area code 212, while the rest of the city became area code 718.

The above text pretty well describes the situation up until the mid 1990's, at which point things changed. With the advent of fax machines, the Internet, and a number of competitive pressures, there was a sudden increase in the demand for telephone service. Both business and residential users wanted more telephones, and more telephone lines. Unfortunately, there were simply not enough area codes and exchanges to go around. This resulted in a number of changes.

NANP Changes

The whole concept of the NANP was modified to allow a much wider range of numbers for both area codes and exchanges. The only rule that still remained of the initial three was that the first digit of an area code or an exchange could be anything except 1 or 0.

This means that numbers like 888 or 890 can now be either area codes or exchange codes. In order to tell what you mean by dialing such a code, the switching equipment now has to count the digits dialed. Here are a few examples:

- 345-6789 is a local call
- 0-345-6789 is also a local call, but it requests the services of an operator, such as to make it a collect call.
- 1-345-678-9876 is a toll call to area code 345, with some prearranged method of paying for the call.
- 0-345-678-9876 is also a toll call to area code 345, but it requests the services of an operator, most likely to arrange payment.
- 0 without any digits following it is a call to the operator.

An interesting side-effect of this system is that there is sometimes a delay before a connection is made. For example, years ago dialing 0 would almost immediately connect you to an operator. Now dialing 0 does nothing for a few seconds, because the switching equipment must wait to see whether you will continue with more digits.

Splits and Overlays

When New York City became too big for a single area code, it was simply split — one part remained area code 212, and the other became new area code 718. The same process was used in other places as well. For example, the entire state of New Jersey was originally area code 201, but was then split into 201 for northern New Jersey, and 609 for southern New Jersey.

Thus in a *split*, a geographic area was split along some boundary, so that all customers in one part had one area code, and customers in the other part had another area code.

Occasionally, customers object to an area code split when they get a new area code. Friends or customers who used the old area code suddenly have to remember a new code. Worse yet, some of them may have to reprogram equipment with the new area code. Furthermore, near the boundary, calling someone a block or two away may require dialing 10 digits (with an area code) rather than the old 7-digit number.

Hence some area code changes have been overlays, rather than splits. In an *overlay*, an additional area code is overlaid — placed on top of — an existing area. For example, New York City is still split into area codes 212 and 718, but now also has area code 917 on top of the old area codes. For example, if a customer with a 212 number in Manhattan orders a new phone line, the new line may be in area code 917. Likewise, a customer with a 718 number in Brooklyn might also have a second line with a 917 area code. Hence area code 917 is overlaid on top of both 212 and 718.

In some parts of the country, there have been heated arguments between the telephone companies, state regulators, and customers over which is better — splitting or overlaying the area code.

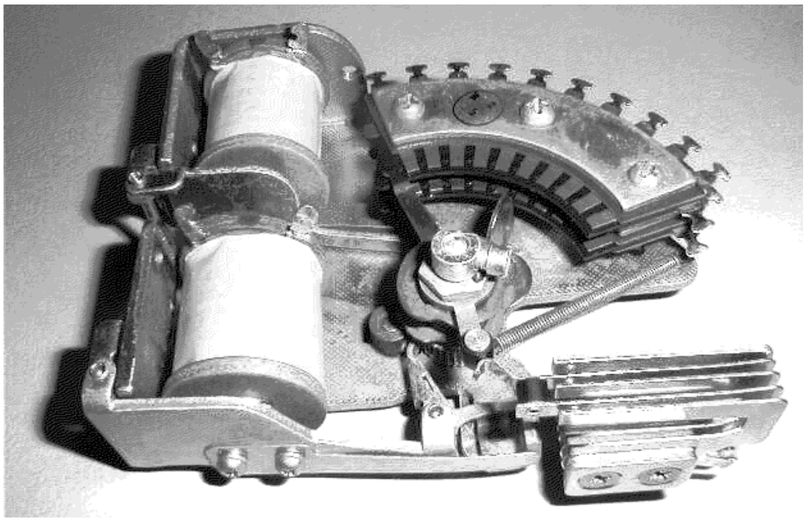

Fig. 11-4. A small stepper relay

Central Office Equipment

The above changes have also been accompanied by major changes in the equipment processing calls in the central office.

The original concept of an exchange, such as PEnnsylvania 6, had all the telephones in the same exchange processed by the same equipment in the central office. In most cases the central office building held the circuitry for one exchange; hence the names *central office* and *exchange* were somewhat synonymous. Only in larger cities might there be two or more sets of exchange equipment housed in the same central office building. Still, an exchange served a specific geographic area.

In the very early days of telephones, the exchange/central office used operators to do the switching. Later, the operators were replaced with electromechanical equipment called a *switch*. Over the years, switches changed technology, but as long as they were electromechanical, they still had one common characteristic — one switch serviced one exchange in one particular area.

So what if you lived in the PEnnsylvania 6 area, but for some reason wanted a GRamercy 5 number? The telephone company could provide such a *foreign exchange* line at extra cost. You would pay a mileage charge for a subscriber loop which would connect your house (near Pennsylvania Station) to the central office switch near Gramercy Park. Your calls would actually be processed by the GRamercy 5 switch, not by the PEnnsylvania 6 switch which serviced the other telephones in your area.

This picture began to change in the 1980's as the telephone network started to become more digitized.

Switches

Although it does not concern us right now, let's list some of the switch functions anyway, just to give a more complete picture:

- It detects and records the number you dial.
- It figures out how to make the connection.
- It connects your call to other parties or circuits.
- It keeps a record of the calls you make.
- It provides extra functions, such as call waiting, call forwarding, or caller ID.

The switch belongs to your local *telco* or telephone company (which may be either the LEC or Local Exchange Carrier, if it is the traditional telco in your area, or a CLEC or Competitive LEC if it is one of the new competitors.)

The original mechanical switches were called *step-by-step* switches because they used a series of circular relays called steppers, which moved or stepped from one contact to the next in direct response to the pulses received from the rotary telephone dial. Fig. 11-4 is a very small stepper relay (much smaller and simpler than the ones that were actually used in telephone switches) that shows the idea of a contact that could move to one of ten output positions. A call would come in via the moving contact, and would then be switched to one of the ten outputs.

Step-by-step switches worked fairly well, but were slow, complicated, and required much maintenance. Electromechanical switches went through a number of changes over the years, and the last such switches were of the *crossbar* type. Fig. 11-5 shows the principle of operation. Imagine a set of vertical contact bars, and a set of horizontal contact bars, separated from each other and not touching. Inputs come in at the left, outputs at the bottom. To make a connection, such as the heavy line in Fig. 11-5, slide the two dark bars toward each other so they touch at the circled point. Now the input signal flows out the appropriate output.

As usual, the actual mechanical crossbar switch was much more complicated than our simple picture. Fig. 11-6 shows just a small section of such a switch. This one has five horizontal bars, each of which could pivot

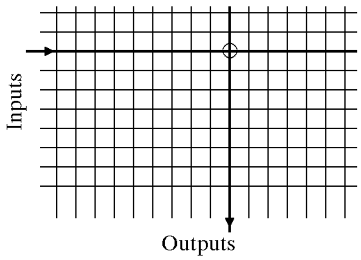

Fig. 11-5. Crossbar switch principle

either up or down, and ten vertical bars which could also move. Mounted at the intersections of the two sets of bars were switch contacts, which would close when the appropriate set of bars moved. This gave a 10×10 matrix of switch contacts which would complete a set of connections as in Fig. 11-5.

During that time, when the switches were all electromechanical, the analog signal on your subscriber loop went directly to the switch, which kept the signal in its analog form. Connections within and between switches generally remained analog, though in some cases the audio signal might be modulated onto a carrier for transmission on a long-distance cable or microwave network.

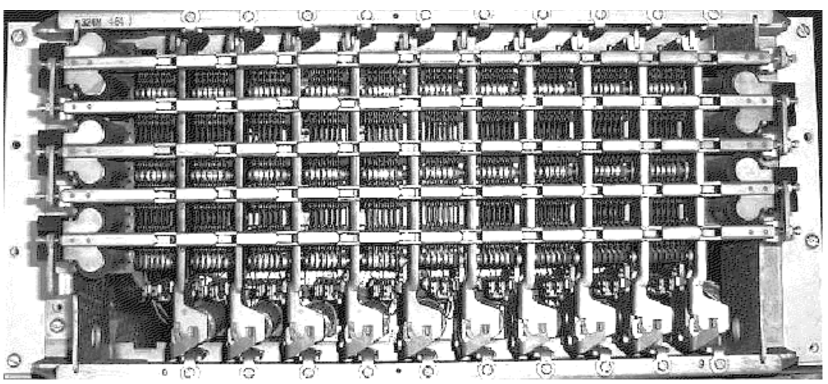

Fig. 11-6. An actual crossbar switch section

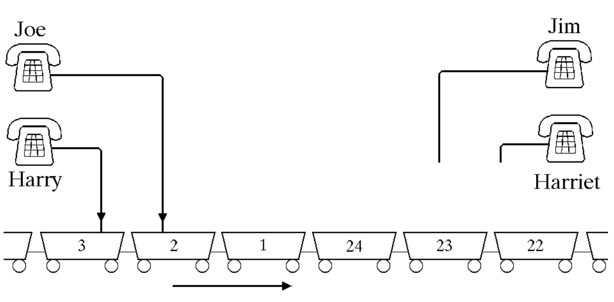

Fig. 11-7. Digital timing slots

Eventually, though, electromechanical switches were replaced with fully electronic switches called Electronic Switching Systems or ESS. Even so, the early ESS switches used digital computers for control, but they still kept the signals in their analog form.

Today, however, things are different. The latest generation of switches is now completely digital. The audio signal from your subscriber loop is converted to a digital signal as soon as it enters the switch. In fact, if part of the connection from your house to the switch is through an optical fiber, then the conversion from analog to digital is done even before the signal reaches the switch. Rather than actually make and break specific connections to set up calls (via stepper switches, crossbar switches, or even simple relays), these digital switches work on an entirely different principle. Fig. 11-7 shows the general idea, called *timing slot assignment.*

Suppose that John on the left wants to talk to Jim on the right, and Harry on the left wants to speak with Harriet on the right. The old analog approach would connect two physical circuits, one between John and Jim, and the other between Harry and Harriet.

Instead, the digital approach sets up one circuit, which all four users (and others as well) share at different times. Imagine that there is a model railroad track running left-to-right across the bottom of Fig. 11-7. There is an infinitely long, never-ending train on it, with little freight cars neatly numbered 1 through 24; the numbers repeat, so after car 24 comes another car 1 etc. John digitizes his voice into binary bits, and each time a car number 2 comes by, he dumps the current load of bits into it. Then he waits until, 23 cars later, another car 2 comes by, and so on. Harry does the same, but with car 3 instead of 2.

Over on the right hand side, we simply tell Jim that his signal is in cars numbered 2, and tell Harriet to wait for cars with the number 3. There is only one track running across the page, but both talkers get to send their voice bits across.

Eventually John and Jim hang up, at which time cars 2 may run empty for a while. Likewise, when Harry and Harriet hang up, their cars numbered 3 become empty, available for someone else.

Let's say that at some later time Harry calls Jim. At the time when he finishes dialing, cars 2 and 3 may be busy, but suppose that car 17 is free. In that case, the switch simply tells Harry to put his bits into car 17, and tells Jim to look for his bits in car 17 as well.

With cars numbered 1 through 24, there is enough capacity for 24 simultaneous conversations. But it is possible to put several hundred telephones at the left side of the drawing, and another several hundred at the right, on the theory that not all of these people will want to talk at the same time — if by chance they do, then some of them will get a *circuit busy* signal because there is not enough capacity for all of their signals to get through. The precise study of how many phones can be at each end, and how often their calls get *blocked* — that is, they do not get through — is called *queing theory*, and we will look at it later.

Regardless of switch type, when you make a call, the switch analyzes the number you dial, and passes your call on. What happens next depends on whom you call, and there are several possibilities:

- If you dial someone who is also served by the same switch (typically someone else in your neighborhood), the switch simply connects your call to that line. This is considered a local call.
- If you dial someone in your general area, but who is served by a different switch, your call will simply be passed from one switch to the other via an interoffice connection called a *trunk*. If both of you are served by the same LEC or CLEC, the process is simple. But if you are served by different telcos, then there has to be an interconnection between the two companies. The CLEC will generally have a *point of presence* or POP — the point where it connects to the local LEC — and your signal will be transferred through there.
- If you dial someone in a different area — perhaps the other side of your state, or the other side of the country — then the call becomes a long-distance toll call, and things become more interesting.

Long Distance Calls

It's obviously not possible for every end office in the entire country to have a direct trunk connection to every other end office. Nearby end offices — and especially those having a lot of traffic between them — might have direct connections, but otherwise a more economical system was necessary.

Fifty years ago, the only long distance carrier was AT&T. All long distance calls, regardless of whether your local telephone company was a member of the Bell System (which was owned by AT&T) or whether it was an independent, were routed through AT&T, and the AT&T network was closely linked to the local telephone system.

Initially, long distance calls were handled manually by operators. When *Direct Distance Dialing* or DDD became available around 1960, the various switches in the overall system were linked as in Fig. 11-8.

At the bottom of Fig. 11-8, each customer's telephone is connected to his local central office, also called a Class 5 office. There are several tens of thousands of such central offices around the country. Nearby central offices are connected by trunks, as shown by the black line. If a call cannot be routed directly to a nearby CO, it would be sent up the line to a Class 4 toll center, which would serve a group of central offices; the toll center would then send the call up the line to a Class 3 primary center.

Nearby primary centers were also linked by trunks, so if the call could be routed that way, it would go from one primary center to another and then down to the customer. If such a trunk connection did not exist (or if all such trunks were currently busy), the call would be sent up to a Class 2 sectional center or perhaps even a Class 1 regional center, which served a large region of the country. In general, the call would be handled at the lowest possible level, and would be sent up the chain only if a more direct connection was not possible.

All the trunks between the various centers were originally audio links — connections which carried audio signals, both telephone conversations as well as signalling tones. Some of them were direct wire connections, while others may have been through microwaves or other systems, but they all carried only audio signals.

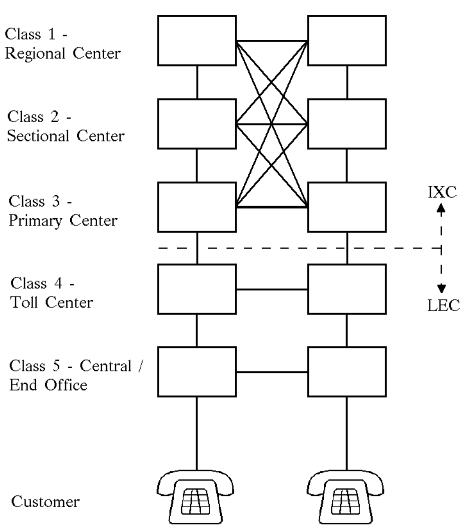

Fig. 11-8. Structure of the long-distance network

If a trunk was free, the two switches at the two ends of the trunk would send a fixed-frequency tone down the line to mark it as free. If one end wanted to use the trunk, it would remove the tone, which would "wake up" the opposite end. It would then send DTMF tones through the trunk to tell the other end what number to connect to. These DTMF tones were similar to the ones used for tone dialing, but the frequencies were different to avoid confusion. Since these signalling tones were carried through the same channel as the phone conversation, this was called *in-band* or *in-channel signalling.*

In-band signalling had several disadvantages:

- Because the signalling tones were carried through the same channel as the phone conversation, some users found ways to duplicate the signalling tones and fool the system into providing free calls. Just for "fun", particularly mischievous users could even control the routing of calls and tie up many trunks by routing the call back and forth through many offices, and even through foreign countries or around the world.
- Each time a call was routed from one switch to the next, the DTMF tones were repeated to set up the next part of the connection. This took up valuable time.
- Even if the called party was busy or did not answer, a complete audio path was still set up from the originating central office to the final one; in fact, that audio path went all the way from the caller's telephone to the telephone being called. Aside again from the possibility of fraud (it was possible to send audio through the chain even if the called party never answered the call), this meant that the telephone system was providing a complete connection for every call, but was not paid for it if the called party was busy or not at home.
- Finally, in-band signalling was very limited. For example, it could send the number of the phone being called, but not the number of the caller. Hence some services which we now take for granted, such as Caller ID, were not possible.

These disadvantages, combined with the advent of digital technology, the breakup (divestiture) of AT&T in 1984, and the rise of competing long-distance carriers, have all led to a major change in the way long distance calls are handled.

SS7 — Signalling System 7

In order to remove signalling from the audio circuits, the telephone companies set up a separate signalling system called Signalling System 7. This is a completely digital system which links the various central offices to centralized computer databases and communications centers. All the signalling is done through this common channel; the process is called CCS or *Common Channel Signalling*, and telephone users have no access to it.

When you place a toll call, the switch in your local central office sends a digital message to a nearby SS7 system database to ask (a) where the called telephone is located and how to connect to it, and (b) whether that phone is free or busy. If the phone number is busy, your local switch almost immediately sends you a busy signal, without even trying to connect the call through.

Aside from reducing the possibility of fraud, having a completely separate digital communications system that handles inter-office communications for telephone switching adds a number of new functions:

- When a call is placed, the SS7 system sends both the calling number and the called number. Since the destination end office knows who placed the call, it can provide Caller ID information.

DETOUR

One of the options with Caller ID is that you can choose to block your number from being displayed. But your number is sent through the SS7 system anyway. If you call an 800 or other free number, your number will be displayed even if you block it, on the theory that the person or company that pays for the call — the recipient of the call — has a right to know who is calling. This version of the system is called ANI or *Automatic Number Identification*.

Companies use this information for various purposes. For example, some mail order companies keep track of the numbers for frequent customers, and will greet you by name as soon as you call. Credit card companies also keep track of your home number, and will only activate a new card if you call from your home number to tell them you received it.

END OF DETOUR

- Since the SS7 database tells the originating end office how to route the call, a call can be routed to a different number than what was dialed. For example, a telephone can have two numbers — a normal number, plus an 800 number. Dialing either number will reach the same phone, but the billing is different. When dialing the normal number, the switch at the caller's end bills the call to the caller. When dialing the 800 number, the switch at the called end bills the call to the recipient.

- Looking up the routing information in an SS7 database also means that routing can change with time or with the place. For example, a nationwide chain of pizza stores can advertise one free 800 number for the entire country. But which local store you get when you dial that number depends on where in the country you are — the call is routed to the store in your area.
- The SS7 system also allows *telephone number portability*, i.e., the ability to take and keep your old phone number when you move out of the area. For example, if you live in New York City and have a telephone number with area code 212, you could *theoretically* move to Los Angeles or Chicago, but keep the same 212 number. We say theoretically because, although this is technically possible, it is not a service that is offered nationally by the telcos because it would tremendously increase the amount of data that has to be stored by the SS7 centralized database computers. Instead, telephone number portability is offered strictly on a local basis. For example, if you switch from your local LEC to a competitive CLEC, you can take your old telephone number with you, even though it is now handled by a different telco and different switch.

LATAs

Local Exchange Carriers — the LECs and CLECs — handle local calls, while the Interexchange Carriers — abbreviated IEC or IXC — handle the long distance calls. In order to mark off which carrier carries what call, the country is divided into LATAs — *Local Access and Transport Areas*. In some cases, a LATA may cover an entire state, while in other cases a state may be divided into several LATAs, or a LATA may stretch into an adjacent state. Furthermore, a LATA may be just one area code, or it may have several area codes, or one area code can span two or three LATAs.

- Example: the New York Metropolitan LATA covers New York City, and roughly 50 miles north and east within New York State, as well as a tiny corner of Connecticut. It contains several area codes, such as 212, 718, 516, 914, and now 845. But part of area code 845 is also in the Poughkeepsie LATA, which is north of New York City.

LECs and CLECs handle calls within a LATA, while calls between LATAs — what are called long-distance calls — are handled by the IXC. Originally the two types of companies were not allowed to compete in each other's turf, but that has changed, as some LECs are now allowed to provide long-distance service as well.

Prices for calls within LATAs (as well as other services) are generally set by the telcos in a document called a *tariff*, which is filed with the public utilities commission of the state that the LATA is in, as are prices for calls between LATAs in the same state. But the prices for interstate calls — that is, for calls between LATAs in different states — are set in tariffs filed with the Federal Communications Commission. An interesting byproduct of this system is that short-distance in-state calls between adjacent LATAs often cost more than interstate calls between two ends of the country.

1-Plus Dialing

In the early days of the North American Numbering Plan (NANP), all area codes had a 0 or 1 as their middle digit; a switch thus recognized a long-distance call by this 0 or 1. When the NANP was changed to allow more area codes and exchange codes, a long-distance call to another area code was marked by dialing a 1 before the area code; this system was called *1-Plus Dialing*.

Originally, however, the only IEC was AT&T, so 1-Plus automatically switched the call to AT&T's facilities. When other long-distance carriers were authorized, another scheme had to be found to allow callers to use them. Eventually, the FCC insisted that they be given equal access to long distance, and so they too were given 1-Plus access. Over time, this system has thus evolved like this:

Initially, AT&T had 1-Plus dialing, and other IECs did not. By default, every telephone user was "pre-subscribed" to AT&T, but could reach another IEC by dialing an 800 number or a 950-xxxx number (such as 950-1022, which was the code for MCI.) This system put the competitive IECs at a disadvantage: since the IEC did not know who the caller was, users had to first open an account with the IEC and get an account number, which they then had to dial. All in all, they generally had to dial eleven digits to reach the IEC, another six to ten digits for an account number, and then ten more digits for the number they wanted to call. On the other hand, this disadvantage was also an advantage — the account could be used from anywhere, including payphones. (Unfortunately, payphone owners then complained that they were not being compensated for the use of their phones for 800-number calls, and the FCC eventually gave them the right to collect a fee from the called party for each free call.)

Later, when 1-Plus dialing was extended to other IECs besides AT&T, each telephone customer had to choose a specific IEC, so that all 1-Plus-dialed calls would be routed to that carrier. It is still, however, possible to use alternate IEC carriers in one of two ways: by opening an account with an IEC and dialing an 800 number (usually by using a prepaid calling card), or by dialing a special 101xxxx code before the 1-Plus number. Both of these alternatives turn out to be more awkward, but quite economical. For example, in the New York area, the code 1010288 selects AT&T, or 1010222 selects MCI. On the other hand, codes such as 1016868 or 1010811 select small, unknown IECs which offer rates significantly lower than the big carriers.

The effect of competition

Carrying a telephone call from one coast to the other is not substantially more expensive than connecting it across town or across a state. Prior to the divestiture of AT&T in 1984, however, long distance calls were priced much higher than local calls. Moreover, business rates were substantially higher than residential rates, and the two in effect subsidized local service.

When competitive companies entered the telephone market after 1984, they naturally went after the high-priced business, where they could offer lower rates and undercut the telephone companies, yet still make a profit. By skimming off the cream of the business — the high-profit areas — they effectively forced the local telephone companies to raise the rates for local service.

As a result, local rates for residential customers have risen substantially over pre-1984 rates. On the other hand, long distance rates have dropped. People who make many long-distance calls come out ahead; people who don't, don't.

One important point is that one must be much more careful to avoid the occasional ripoff. For example, to call central Europe, one company may charge $3 per minute, while another charges 29 cents. Even the same company may charge different rates to different customers. For example, one AT&T customer may pay 25 cents per minute to call coast to coast, while another pays 5.9 cents per minute.

COCOTs — customer owned coin operated telephones — are another example. Most states require that local calls from COCOT telephones cost the same as those from telco-owned payphones. But long-distance calls are not so regulated. For example, a toll call from a telco payphone may simply be routed to a major carrier like AT&T or Sprint, resulting in a rate which, while not exactly cheap, is not wildly unreasonable either. But a toll call from a COCOT phone is often handled differently.

A payphone needs a certain amount of "smarts" to function. Some device must keep track of time, compute charges for calls, count coins, and identify the phone to the IEC on a long-distance call. Telco-owned payphones are generally "dumb" phones, with all the smarts located in the central office switch. They are connected to a special dial tone line, which identifies them as a payphone to the switch.

Many COCOT payphones, on the other hand (especially those from tiny COCOT companies) are often connected to a regular business line. They do not have the help of the central office and its switch. The COCOT phone is therefore a "smart phone" — it has a built-in computer which keeps track of call charges and time, counts coins, and keeps you from making certain calls, such as calling the local operator. Instead, when you pick up the phone, the dial tone you hear is generated locally in the phone. As you dial, the built-in computer monitors your dialing and decides what to do. If it recognizes the number as a local call, a voice synthesizer tells you to insert the appropriate coins. After you do that, the phone will connect to the telephone line, and redial your number to actually place the call.

If the COCOT phone does not recognize a local number, it stores the number you dialed and instead dials an 800 number to reach the COCOT operator, who might be on the other side of the country, or perhaps even in some foreign country where labor rates are cheaper. When that call goes through, the phone repeats the number you dialed, and connects you to the operator, who will now arrange payment and complete the call for you. A simple call from, say, New York to New Jersey, might thus really become two calls — one from New York to an operator who might perhaps be in Texas, and another from Texas back to New Jersey. The result is a much more expensive call.

Unlike telco-owned payphones, COCOT phones also don't generally allow incoming calls. They do not list their own telephone number on the front of the phone, and may not even contain a bell. The reason is simple — the computer inside the telephone frequently needs to be reprogrammed with new instructions, new numbers or new prices. It therefore has an internal modem, which lets the COCOT company call in changes, do remote troubleshooting, or call in to determine whether the coin box needs to be emptied. Since the modem answers incoming calls, they do not want customers to call in as well.

