computers are bad

You are receiving this facsimile because you signed up for fax delivery of this newsletter. To stop delivery, contact Computers Are Bad by email or fax.

https://computer.rip - me@computer.rip - fax: +1 (505) 926-5492

2021-12-05 singing wires

I have probably described before the concept of the telephone network forming a single, continuous pair of wires from your telephone to the telephone of the person you are calling. This is the origin of “circuit switching” and the source of the term: the notion that a circuit-switched system literally forms an electrical circuit between two endpoints.

Of course, we presumably understand that modern systems don’t actually do this. For one, most long-distance data transmission today is by means of optics. And more importantly, most modern systems that we call circuit-switching are really, in implementation, packet switching systems that use fixed allocation schemes to provide deterministic behavior that is “as good as a circuit.”

Consider the case of MPLS, or Multi-Protocol Label Switching, a network protocol which was formerly extremely popular in telecom and ISP backhaul networks and is still common today, although improvements in IP switching have reduced its popularity [1]. MPLS is a “circuit-switched” system in that it establishes “virtual circuits,” i.e. connection setup is a separate operation from using the connection. But, in implementation, MPLS is a packet switching system and inherits the standard limitations thereof (resulting in a need for QoS and traffic engineering mechanisms to provide deterministic performance). We can say that MPLS implements circuit switching on top of packet switching. One of the fun things about networking is that we can do things like this.

Why, though? Circuit switching is, conceptually, very simple. So why do we bother with things like MPLS that make it very much more complicated, even as simple as MPLS is?

There are two major reasons, one fundamental and one practical. First, the conventional naive explanation of circuit switching implies that, when I call someone in India, the telephone network allocates a set of copper wires all the way from me to them. This is a distance of many thousands of miles, which includes oceans, and it does not seem especially likely that the telecom industry has sunk the thousands of tons of copper into the ocean that would be required to accommodate the telephone traffic between the US and the Asian region. It is obvious on any consideration that, somehow, my telephone call is being combined with other telephone calls onto a shared medium.

Second, there is the issue of range. The microwatt signal produced by your telephone will not endure thousands of miles of wire, even if the gauge was made unreasonably large. For this simple practical reason, signals being moved over long distances need to somehow be encoded differently in a way that can cover very long distances.

Both of these things are quite unsurprising to us today, because we are fortunate enough to live in a world in which these problems were solved long ago. Today, I’m going to talk about how these problems were solved, in the first case where they were encountered at large scale: the telephone network.

Your landline telephone is connected by means of a single pair, two wires. This pair forms a *loop,* called the local loop, from the exchange to your phone and back. Signals are conveyed by varying the voltage (and, by the same token, current as the resistance is fixed) on the circuit, or in other words by amplitude modulation. The fact that this works in full duplex on a single loop is surprisingly clever from the modern perspective of digital protocols which almost universally are either half-duplex or need separate paths for each direction, but the electrical trick that enables this was invented at about the same time as the telephone. It’s reasonably intuitive, although not quite technically accurate, to say that each end of the telephone line knows what signal it is sending and can thus subtract it from the line potential.

The possible length of the local loop is limited. It varies by the gauge of wire used, which telephone companies selected based on loop length to minimize their costs. In general, beyond about ten miles the practicality starts to drop as more and more things need to be done to the line to adjust for resistance and inductance attenuating the signal.

The end result is that the local loop, the part of the telephone system we are used to seeing, is actually sort of the odd one out. Virtually every other part of the telephone system uses significantly different signaling methods to convey calls, and that’s not just a result of digitization: it’s pretty much always been that way, since the advent of long-distance telephony.

Before we get too much further into this, though, a brief recap of the logical architecture of the telephone system. Let’s say you make a long distance call.

In simplified form (in the modern world there are often more steps for optimization reasons), your phone is directly connected by the local loop to a *class 5* or *local* switch in your exchange office. The local switch consults routing information and determines that the call cannot be completed locally, so it connects your call to a *trunk*. A trunk is a phone line that does not connect a switch to a phone... instead, it connects a switch to another switch. Trunk lines thus make up the backbone of the telephone network.

In this case, the trunk line will go to a *tandem*, *class 4*, or *toll* switch. These are all mostly interchangeable terms used at different periods. Tandem switches, like trunk lines, are not connected to any subscriber phones. Their purpose is to route calls from switch to switch, primarily to enable long distance calling--between two local switches. In our example, the tandem switch may either select a trunk to the local switch of the called party, or if there is no one-hop route available it will select a trunk to another tandem switch which is closer to [2] the called party. Eventually, the last tandem switch in the chain will select a trunk line to the called party’s local switch, which will select the local loop to their phone.

What we are most interested in, here, are the trunks.

Trunk lines may be very long, reaching thousands of miles for trans-continental calls. They are also expected to serve a high capacity. Almost regardless of the technology [3], laying new trunk lines is a considerable expense to this day, so it’s desirable to concentrate a very large amount of traffic onto a small number of major lines. As a result, common routing in the telephone network tends to resemble a hub and spoke architecture, with calls between increasingly larger regions being concentrated onto just a few main trunks between those regions. The modern more mesh-like architecture of the internet, the more flexible routing technology it required, and the convergence of telephony on IP is lessening this effect, but it's still fairly prominent and was completely true of the early long-distance network.

Consider, for example, calls from New York City to Los Angeles. These two major cities are separated by a vast distance, yet many calls are placed between them. For cost reasons, just a small number of transcontinental lines, each of them a feat of engineering, must take the traffic. Onto those same lines is aggregated basically the entire call volume between the east coast and the west coast, easily surpassing one hundred thousand simultaneous connections.

Now, imagine you tried to do this by stringing one telephone line for each call.

Well, in the earliest days of long-distance telephony, that's exactly what was done. A very long two-wire telephone circuit was strung between cities just like between exchange offices and homes. To manage the length, inductance coils were added at frequent intervals to adjust frequency response, and at less frequent intervals the line was converted to four-wire (one pair each direction) so that an amplifier could be inserted in each pair to "boost" the signal against line loss. These lines were expensive and the quality of the connection was poor, irritating callers with low volume levels and excessive noise.

Very quickly, long-distance trunks were converted to a method we now refer to as *open wire*. On these open wire trunks, sets of four wires (one pair for each direction) were strung alongside each other across poles with multiple cross-arms. Because a set of four wires was required for every simultaneous phone call the trunk could support (called a channel), it was common to have an absolute maze of wires as, say, four cross-arms on each pole each supported two four-wire pairs. This large, costly assembly supported only eight channels.

Four-wire circuits were used instead of two-wire circuits for several reasons, including lower loss and greater noise immunity. But moreover, continuous use of a four-wire circuit made it easier to install amplifiers without having to convert back and forth (which somewhat degraded quality every time). Loading coils to adjust inductance were still installed at regular intervals (every mile was typical).

The size and cost of these trunks was huge. Nonetheless, in 1914 AT&T completed the first transcontinental telephone trunk, connecting for the first time the eastern network (through Denver) to the previously isolated west coast network. The trunk used three amplifiers and uncountable loading coils. Amusingly, for basically marketing reasons, it would not go into regular service until 1915.

The high cost of this and subsequent long-distance connections was a major contributor to the extraordinary cost of long-distance calls, but demand was high and so long-distance open-wire trunks were extensively built, especially in the more densely populated northeast where they formed the primary connections between smaller cities for decades to come.

Years later, the development of durable, low-cost plastics considerably reduced the cost of these types of trunks by enabling cheap "sheathed" cables. These cables combined a great number of wire pairs into a single, thick cable that was far cheaper and faster to install over long distances. Nonetheless, the fundamental problem of needing two pairs for each channel and extensive line conditioning remained much the same. The only real difference in call quality was that sheathed cables avoided the problem of partially shorting due to rain or snow, which used to make open-wire routes very poor during storms.

It was clear to Bell System engineers that they needed some form of what we now call multiplexing: the ability to place multiple phone calls onto a single set of wires. The first, limited method of doing so was basically a way of intentionally harnessing crosstalk, the tendency of signals on one pair to “leak” onto the pair run next to it. By use of a clever transformer arrangement, two pairs could each carry one direction of one call... and the two pairs together, each used as one wire of a so-called phantom circuit, could carry one direction of a third call. This represented a 50% increase in capacity, and the method was widely used on inter-city trunks. Unfortunately, combining phantom circuits into additional super-phantom circuits proved impractical, and so the modest 1.5x improvement remained and the technique was far from addressing the problem.

Vacuum tube technology, originally employed for amplifiers on open-wire circuits, soon offered an interesting new potential: carriers. Prior to carrier methods, all telephone calls were carried on trunks in the audible frequency range, just like on local loops. Carrier systems entailed using the audio frequency signal to modulate a higher frequency carrier, much like radio. At the other end, the carrier frequency could be isolated and the original audio frequency demodulated. By mixing multiple carriers together, multiple channels could be placed on the same open-wire pair with what we now call frequency-division muxing.

The first such multiplexed trunk went into service in 1918 using what AT&T labeled the “A” carrier. A-carrier was capable of carrying four channels on a pair, using a single-sideband signal with suppressed carrier frequency, much like the radio systems of the time. These carrier systems operated above audible frequency (voice frequency or VF) and so were not considered to include the VF signal, with the result that an open-wire line with A carrier could convey five channels: four A-carrier channels and one VF channel.

Subsequent carriers were designed to use FDM on both open-wire and sheathed cables, using improved electronics to fit more channels. Further, carriers could be used to isolate the two directions instead of separate pairs, once again allowing full-duplex operation on a single wire pair while still keeping amplifiers practical.

This line of development culminated in the J-carrier, which placed 12 channels on a single open-wire trunk. J-carrier operated above the frequencies used by older carriers such as C-carrier and VF, and so these carriers could be “stacked” to a degree enabling a total of 17 bidirectional channels on a four-wire trunk [4], using frequencies up to 140 KHz. This 17x improvement came at the cost of relatively complex electronics and more frequent amplifiers, but still yielded a substantial cost reduction on a per-channel basis. J-carrier was widely installed in the 1920s as an upgrade to existing open-wire trunks.

Sheathed cables yielded somewhat different requirements, as crosstalk was a greater issue. A few methods of mitigating the problem lead to the development of the K-carrier, which multiplexed 12 channels onto each pair in a sheathed cable. Typically, one sheathed cable was used for signals each direction to reduce crosstalk. Sheathed cables could contain a large number of pairs (hundreds was typical), making the capacity highly scalable. Further, K-carrier was explicitly designed to operate without loading coils, further lessening cost of the cable itself. In fact, loading coils improved frequency response only to a point and worsened it much beyond VF, so later technologies like K-carrier and even DSL required that any loading coils on the line be removed.

As a downside, K-carrier required frequent repeaters: every 17 miles. Each repeater consisted of two amplifiers, one each direction, per pair in use. Clever techniques which I will not describe in depth were used to automatically adjust amplifiers to maintain consistent signal levels throughout the line. Because these repeaters were fairly large and power intensive, they were installed in fairly substantial brick buildings that resembled small houses but for their unusual locations. Three-phase power had to be delivered to each building, usually adding additional poles and wires.

The size of the buildings is really quite surprising, but we must remember that this was still prior to the invention of the transistor and so the work was being done by relatively large, low-efficiency tubes, with sensitive environmental requirements. The latter was a particularly tricky aspect of analog carriers. Repeater buildings for most open-wire and cable carriers used extremely thick brick walls, which was not yet for blast hardening but instead a method of passive temperature stabilization as the thermal mass of the brick greatly smoothed the diurnal temperature cycle. A notable K-carrier trunk ran between Denver and El Paso, and the red brick repeater buildings can still be seen in some more rural places from I-25.

This post has already reached a much greater length than I expected, and I have yet to reach the topics that I intended to spend most of it on (coaxial and microwave carriers). So, let's call this Part I, and look forward to Part II in which the telephone network will fight the Cold War.

[1] MPLS used to have massive dominance because it was practical to implement MPLS switching in hardware, and IP switching required software. Of course, the hardware improved and IP switching can now be done in silicon, which reduces the performance advantage of MPLS. That said, MPLS continues to have benefits and new MPLS systems are still being installed.

[2] Closer in the sense of network topology, not physical locality. The topology of the telephone network often reflects history and convenience, and so can have unexpected results. For much of the late 20th century, virtually all calls in and out of the state of New Mexico passed through Phoenix, even if they were to or from Texas. This was simply because the largest capacity trunk out of the state was a fiber line from the Albuquerque Main tandem to the Phoenix Main tandem, along the side of I-40. Phoenix, being by then a more populous city, was better connected to other major cities.

[3] Basically the only exception is satellite, for which the lack of wires means that cost tends to scale more with capacity than with distance. But geosynchronous satellites introduce around a half second of latency, which telephone callers absolutely despised. AT&T's experiments with using satellites to connect domestic calls were quickly abandoned due to customer complaints. Satellites are avoided even for international calls, with undersea cables much preferred in terms of customer experience. Overall the involvement of satellites in the telephone network has always been surprisingly minimal, with their role always basically limited to connecting small or remote countries to which there was not yet sufficient cable capacity.

[4] Because the VF or non-carrier channel could be used by any plain old telephone connected to the line (via a hybrid transformer to two-wire), it was used as an *order*

wire. The order wire was essentially a “bonus” channel that was primarily used by linemen to communicate with exchange office staff during field work, i.e. to obtain their orders. While radio technology somewhat obsoleted this use of the order wire, it remained useful for testing and for connecting automated maintenance alarms. Telephone carriers to this day usually have some kind of dedicated order wire feature.