computers are bad

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

2024-02-25 a history of the tty

It's one of those anachronisms that is deeply embedded in modern technology. From cloud operator servers to embedded controllers in appliances, there must be uncountable devices that think they are connected to a TTY.

I will omit the many interesting details of the Linux terminal infrastructure here, as it could easily fill its own article. But most Linux users are at least peripherally aware that the kernel tends to identify both serial devices and terminals as TTYs, assigning them filesystem names in the form of /dev/tty*. Probably a lot of those people remember that this stands for teletype or perhaps teletypewriter, although in practice the term teleprinter is more common.

Indeed, from about the 1950s (the genesis of electronic computers) to the 1970s (the rise of video display terminals/VDTs), teleprinters were the most common form of interactive human-machine interface. The "interactive" distinction here is important; early computers were built primarily around noninteractive input and output, often using punched paper tape. Interactive operation was a more advanced form of computing, one that took almost until the widespread use of VDTs to mature. Look into the computers of the 1960s especially, the early days of interactive operation, and you will be amazed at how bizarre and unfriendly the command interface is. It wasn't really intended for people to use; it was for the Computer Operator (who had attended a lengthy training course on the topic) to troubleshoot problems in the noninteractive workload.

But interactive computing is yet another topic I will one day take on. Right now, I want to talk about the heritage of these input/output mechanisms. Why is it that punched paper tape and the teleprinter were the most obvious way to interact with the first electronic computers? As you might suspect, the arrangement was one of convenience. Paper tape punches and readers were already being manufactured, as were teleprinters. They were both used for communications.

Most people who hear about the telegraph think of Morse code keys and rhythmic beeping. Indeed, Samuel Morse is an important figure in the history of telegraphy. The form of "morse code" that we tend to imagine, though, a continuous wave "beep," is mostly an artifact of radio. For telegraphs, no carrier wave or radio modulation was required. You can transmit a message simply by interrupting the current on a wire.

This idea is rather simple to conceive and even to implement, so it's no surprise that telegraphy has a long history. By the end of the 18th century inventors in Europe and Great Britain were devising simple electrical telegraphs. These early telegraphs had limited ranges and even more limited speeds, though, a result mostly of the lack of a good way to indicate to the operator whether or not a current was present. It is an intriguing aspect of technical history that the first decades of experimentation with electricity were done with only the clumsiest means of measuring or even detecting it.

In 1820, three physicists or inventors (these were vague titles at the time) almost simultaneously worked out that electrical current induced a magnetic field. They invented various ways of demonstrating the effect, usually by deflecting a magnetic needle. This innovation quickly lead to the "electromagnetic telegraph," in which a telegrapher operates a key to switch current, which causes a needle or flag to deflect at the other end of the circuit. This was tremendously simpler than previous means of indicating current and was applied almost immediately to build the first practical telegraphs. During the 1830s, the invention of the relay allowed telegraph signals to be repeated or amplified as the potential weakened (the origin of the term "relay"). Edward Davy, one of the inventors of the relay, also invented the telegraph recorder.

From 1830 to 1850, so many people invented so many telegraph systems that it is difficult to succinctly describe how an early practical telegraph worked. There were certain themes: for non-recording systems, a needle was often deflected one way or the other by the presence or absence of current, or perhaps by polarity reversal. Sometimes the receiver would strike a bell or sound a buzzer with each change. In recording systems, a telegraph printer or telegraph recorder embossed a hole or left a small mark on a paper tape that advanced through the device. In the first case, the receiving operator would watch the needle, interpreting messages as they came. In the second case, the operator could examine the paper tape at their leisure, interpreting the message based on the distances between the dots.

Recording systems tended to be used for less time-sensitive operations like passing telegrams between cities, while non-recording telegraphs were used for more real-time applications like railroad dispatch and signaling. Regardless, it is important to understand that the teleprinter is about as old as the telegraph. Many early telegraphs recorded received signals onto paper.

The interpretation of telegraph signals was as varied as the equipment that carried them. Samuel Morse popularized the telegraph in the United States based in part on his alphabetic code, but it was not the first. Gauss famously devised a binary encoding for alphabetic characters a few years earlier, which resembles modern character encodings more than Morse's scheme. In many telegraph applications, though, there was no alphabetic code at all. Railroad signal telegraphs, for example, often used application-specific schemes that encoded types of trains and routes instead of letters.

Morse's telegraph system was very successful in the United States, and in 1861 a Morse telegraph line connected the coasts. It surprises some that a transcontinental telegraph line was completed some fifty years before the transcontinental telephone line. Telegraphy is older, though, because it is simpler. There is no analog signaling involved; simple on/off or polarity signals can be amplified using simple mechanical relays. The tendency to view text as more complex than voice (SMS came after the first cellphones, for one) has more to do with the last 50 years than the 50 years before.

The Morse telegraph system was practical enough to spawn a large industry, but suffered a key limitation: the level of experience required to key and copy Morse quickly and reliably is fairly high. Telegraphers were skilled and, thus, fairly well paid and sometimes in short supply [1]. To drive down the cost of telegraphy, there would need to be more automation.

Many of the earliest telegraph designs had employed parallel signaling. A common scheme was to provide one wire for each letter, and a common return. These were impractical to build over any meaningful distance, and Morse's one-wire design (along with one-wire designs by others) won out for obvious reasons. The idea of parallel signaling stayed around, though, and was reintroduced during the 1840s with a simple form of multiplexing: one "logical channel" for each letter could be combined onto one wire using time division muxing, for example by using a transmitter and receiver with synchronized spinning wheels. Letters would be presented by positions on the wheel, and a pulse sent at the appropriate point in the revolution to cause the teleprinter to produce that letter. With this alphabetic teleprinter, an experienced operator was no longer required to receive messages. They appeared as text on a strip of paper, ready for an unskilled clerk to read or paste onto a message card.

This system proved expensive but still practical to operate, and a network of such alphabetic teleprinters was built in the United States during the mid 19th century. A set of smaller telegraph companies operating one such system, called the Hughes system after its inventor, joined together to become the Western Union Telegraph Company. In a precedent that would be followed even more closely by the telephone system, practical commercial telegraphy was intertwined with a monopoly.

The Hughes system was functional but costly. The basic idea of multiplexing across 30 channels was difficult to achieve with mechanical technology. Émile Baudot was employed by the French telegraph service to find a way to better utilize telegraph lines. He first developed a proper form of multiplexing, using synchronized switches to combine five Hughes system messages onto one wire and separate them again at the other end. Likely inspired by his close inspection of the Hughes system and its limitations, Baudot went on to develop a more efficient scheme for the transmission of alphabetic messages: the Baudot code.

Baudot's system was similar to the Hughes system in that it relied on a transmitter and receiver kept in synchronization to interpret pulses as belonging to the correct logical channel. He simplified the design, though, by allowing for only five logical channels. Instead of each pulse representing a letter, the combination of all five channels would be used to form one symbol. The Baudot code was a five-bit binary alphabetic encoding, and most computer alphabetic encodings to the present day are at least partially derived from it.

One of the downsides of Baudot's design is that it was not quite as easy to operate as telegraphy companies would hope. Baudot equipment could keep up 30 words per minute with a skilled operator who could work the five-key piano-style keyboard in good synchronization with the mechanical armature that read it out. This took a great deal of practice, though, and pressing keys out of synchronization with the transmitter could easily cause incorrect letters to be sent.

In 1901, during the early days of the telephone, Donald Murray developed an important enhancement to the Baudot system. He was likely informed by an older practice that had been developed for Morse telegraphs, of having an operator punch a Morse message into paper tape to be transmitted by a simple tape reader later. He did the same for Baudot code: he designed a device with an easy to use typewriter-like keyboard that punched Baudot code onto a strip of paper tape with five rows, one for each bit. The tape punch had no need to be synchronized with the other end, and the operator could type at whatever pace they were comfortable.

The invention of Murray's tape punch brought about the low-cost telegram networks that we are familiar with from the early 20th century. A clerk would take down a message and then punch it onto paper tape. Later, the paper tape would be inserted into a reader that transmitted the Baudot message in perfect synchronization with the receiver, a teleprinter that typed it onto tape as text once again. The process of encoding and decoding messages for the telegraph was now fully automated.

The total operation of the system, though, was not. For one, the output was paper tape, that had to be cut and pasted to compose a paragraph of text. For another, the transmitting and receiving equipment operated continuously, requiring operators to coordinate on the scheduling of sending messages (or they would tie up the line and waste a lot of paper tape).

In a wonderful time capsule of early 20th century industrialism, the next major evolution would come about with considerable help from the Morton Salt Company. Joy Morton, its founder, agreed to fund Frank Pearne's efforts to develop an even more practical printing telegraph. This device would use a typewriter mechanism to produce the output as normal text on a page, saving considerable effort by clerks. Even better, it would use a system of control codes to indicate the beginning and end of messages, allowing a teleprinter to operate largely unattended. This was more complex than it sounded, as it required finding a way for the two ends to establish clock synchronization before the message.

There were, it turned out, others working on the same concept. After a series of patent disputes, mergers, and negotiations, the Morkrum-Kleinschmidt Company would market this new technology. A fully automated teleprinter, lurching into life when the other end had a message to send, producing pages of text like a typewriter with an invisible typist.

In 1928, Morkrum-Kleinschmidt adopted a rather more memorable name: the Teletype Corporation. During the development of the Teletype system, the telephone network had grown into a nationwide enterprise and one of the United States' largest industrial ventures (at many points in time, the country's single largest employer). AT&T had already entered the telegraph business by leasing its lines for telegraph use, and work had already begun on telegraphs that could operate over switched telephone lines, transmitting text as if it were a phone call. The telephone was born of the telegraph but came to consume it. In 1930, the Teletype Corporation was purchased by AT&T and became part of Western Electric.

That same year, Western Electric introduced the Teletype Model 15. Receiving Baudot at 45 baud [2] with an optional tape punch and tape reader, the Model 15 became a workhorse of American communications. By some accounts, the Model 15 was instrumental in the prosecution of World War II. The War Department made extensive use of AT&T-furnished teletype networks and Model 15 teleprinters as the core of the military logistics enterprise. The Model 15 was still being manufactured as late as 1963, a production record rivaled by few other electrical devices.

It is difficult to summarize the history of the networks that teleprinters enabled. The concept of switching connections between teleprinters, as was done on the phone network, was an obvious one. The dominant switched teleprinter network was Telex, not really an organization but actually a set of standards promulgated by the ITU. The most prominent US implementation of Telex was an AT&T service called TWX, short for Teletypewriter Exchange Service. TWX used Teletype teleprinters on phone lines (in a special class of service), and was a very popular service for business use from the '40s to the '70s.

Incidentally, TWX was assigned the special purpose area codes 510, 610, 710, 810, and 910, which contained only teleprinters. These area codes would eventually be assigned to other uses, but for a long time ranked among the "unusual" NPAs.

Western Union continued to develop their telegraph network during the era of TWX, acting in many ways as a sibling or shadow of AT&T. Like AT&T, Western Union developed multiplexing schemes to make better use of their long-distance telegraph lines. Like AT&T, Western Union developed automatic switching systems to decrease operator expenses. Like AT&T, Western Union built out a microwave network to increase the capacity of their long-haul network. Telegraphy is one of the areas where AT&T struggled despite their vast network, and Western Union kept ahead of them, purchasing the TWX service from AT&T. Western Union would continue to operate the switched teleprinter network, under the Telex name, into the '80s when it largely died out in favor of the newly developed fax machine.

During the era of TWX, encoding schemes changed several times as AT&T and Western Union developed better and faster equipment (Western Union continued to make use of Western Electric-built Teletype machines among other equipment). ASCII came to replace Baudot, and so a number of ASCII teleprinters existed. There were also hybrids. For some time Western Union operated teleprinters on an ASCII variant that provided only upper case letters and some punctuation, with the benefit of requiring fewer bits. The encoding and decoding of this reduced ASCII set was implemented by the Bell 101 telephone modem, designed in 1958 to allow SAGE computers to communicate with one another and then widely included in TWX and Telex teleprinters. The Bell 101's descendants would bring about remote access to time-sharing computer systems and, ultimately, one of the major forms of long-distance computer networking.

You can see, then, that the history of teleprinters and the history of computers are naturally interleaved. From an early stage, computers operated primarily on streams of characters. This basic concept is still the core of many modern computer systems and, not coincidentally, also describes the operation of teleprinters.

When electronic computers were under development in the 1950s and 1960s, teleprinters were near the apex of their popularity as a medium for business communications. Most people working on computers probably had experience with teleprinters; most organizations working on computers already had a number of teleprinters installed. It was quite natural that teleprinter technology would be repurposed as a means of input and output for computers.

Some of the very earliest computers, for example those of Konrad Zuse, employed punched tape as an input medium. These were almost invariably repurposed or modified telegraphic punched tape systems, often in five-bit Baudot. Particularly in retrospect, as more materials have become available to historians, it is clear that much of the groundwork for digital computing was laid by WWII cryptological efforts.

Newly devised cryptographic machines like the Lorenz ciphers were essentially teleprinters with added digital logic. The machines built to attack these codes, like Colossus, are now generally recognized as the first programmable computers. The line between teleprinter and computer was not always clear. As more encoding and control logic was added, teleprinters came to resemble simple computers.

The Manchester Mark I, a pioneer of stored-program computing built in 1949, used a 5-bit code adopted from Baudot by none other than Alan Turing. The major advantage of this 5-bit encoding was, of course, that programs could be read and written using Baudot tape and standard telegraph equipment. The addition of a teleprinter allowed operators to "interactively" enter instructions into the computer and read the output, although the concept of a shell (or any other designed user interface) had not yet been developed. EDSAC, a contemporary of the Mark I and precursor to a powerful tea logistics system that would set off the development of business computing, also used a teleprinter for input and output.

Many early commercial computers limited input and output to paper tape, often 5-bit for Baudot or 8-bit for ASCII with parity, as in the early days of computing preparation of a program was an exacting process that would not typically be done "on the fly" at a keyboard. It was, of course, convenient that teleprinters with tape punches could be used to prepare programs for entry into the computer.

Business computing is most obviously associated with IBM, a company that had large divisions building both computers and typewriters. The marriage of the two was inevitable considering the existing precedent. Beginning around 1960 it was standard for IBM computers to furnish a teleprinter as the operator interface, but IBM had a distinct heritage from the telecommunications industry and, for several reasons, was intent on maintaining that distinction. IBM's teleprinter-like devices were variously called Data Communications Systems, Printer-Keyboards, Consoles, and eventually Terminals. They generally operated over proprietary serial channels.

Other computer manufacturers didn't have typewriter divisions, and typewriters and teleprinters were actually rather complex mechanical devices and not all that easy to build. As a result, they tended to buy teleprinters from established manufacturers, often IBM or Western Electric. Consider the case of a rather famous non-IBM computer, the DEC PDP-1 of 1960. It came with a CRT graphics display as standard, and many sources will act as if this was the primary operator interface, but it is important to understand that early CRT graphics displays had a hard time with text. Text is rather complex to render when you are writing point-by-point to a CRT vector display from a rather slow machine. You would be surprised how many vertices a sentence has in it.

So despite the ready availability of CRTs in the 1960s (they were, of course, well established in the television industry), few computers used them for primary text input/output. Instead, the PDP-1 was furnished with a modified IBM typewriter as its console. This scheme of paying a third-party company (Soroban Engineering) to modify IBM typewriters for teleprinter control was apparently not very practical, and later DEC PDP models tended to use Western Electric Teletypes as user terminals. These had the considerable advantage that they were already designed to operate over long telephone circuits, making it easy to install multiple terminals throughout a building for time sharing use.

Indeed, time sharing was a natural fit for teleprinter terminals. With a teleprinter and a computer with a suitable modem, you could "call in" to a time sharing computer over the telephone from a remote office. Most of the first practical "computer networks" (term used broadly) were not actually networks of computers, but a single computer with many remote terminals. This architecture evolved into the BBS and early Internet-like services such as CompuServe. The idea was surprisingly easy to implement once time sharing operating systems were developed; the necessary hardware was already available from Western Electric.

While I cannot swear to the accuracy of this attribution, many sources suggest that the term "tty" as a generic reference to a user terminal or serial I/O channel originated with DEC. It seems reasonable; DEC's software was very influential on the broader computer industry, particularly outside of IBM. UNIX originally targeted a PDP-11 with teleprinters. While I can't prove it, it seems quite believable that the tty terminology was adopted directly from RT-11 or another operating system that Bell Labs staff might have used on the PDP-11.

Computers were born of the teleprinter and would inevitably come to consume them. After all, what is a computer but a complex teleprinter? Today, displaying text and accepting it from a keyboard is among the most basic functions of computers, and computers continue to perform this task using an architecture that would be familiar to engineers in the 1970s. They would likely be more surprised by what hasn't changed than what has: many of us still spend a lot of time in graphical software pretending to be a video display terminal built for compatibility with teleprinters.

And we're still using that 7-bit ASCII code a lot, aren't we. At least Baudot died out and we get to enjoy lower case letters.

[1] Actor, singer, etc. Gene Autry had worked as a telegrapher before he began his career in entertainment. This resulted in no small number of stories of a celebrity stand-in at the telegraph office. Yes, this is about to be a local history anecdote. It is fairly reliably reported that Gene Autry once volunteered to stand in for the telegrapher and station manager at the small Santa Fe Railroad station in Socorro, New Mexico, as the telegrapher had been temporarily overwhelmed by the simultaneous arrival of a packed train and a series of telegrams. There are enough of these stories about Gene that I think he really did keep his Morse sharp well into his acting career.

[2] Baud is a somewhat confusing unit derived from Baudot. Baud refers to the number of symbols per second on the underlying communication medium. For simple binary systems (and thus many computer communications systems we encounter daily), baud rate is equivalent to bit rate (bps). For systems that employ multi-level signaling, the bit rate will be higher than the baud rate, as multiple bits are represented per symbol on the wire. Methods like QAM are useful because they result in bit rates that are many multiples of the baud rate, reducing the bandwidth on the wire.