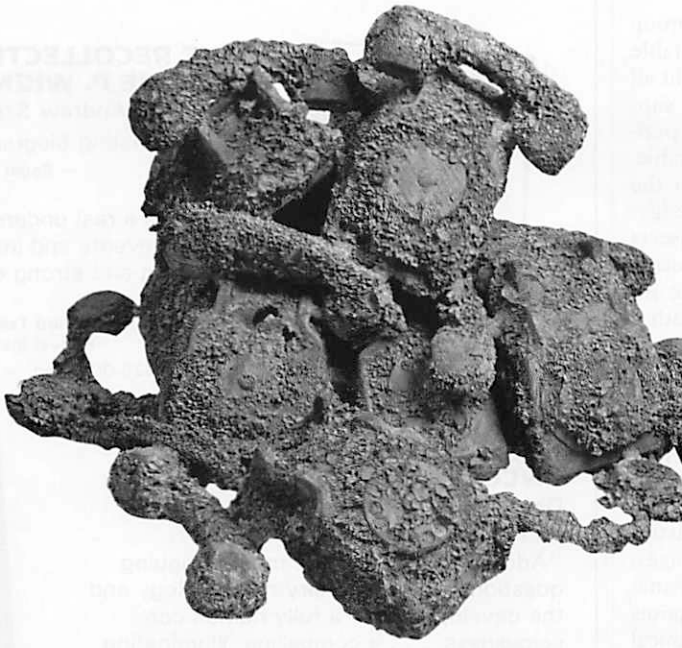
Brian Hayes



Arman, Octopus, *1990*

The Numbering Crisis in World Zone 1

Scarcity is no stranger in this land of plenty. From time to time it seems we are running out of fuel, out of water, out of housing, out of wilderness, out of ozone, out of places to put the rubbish, out of all the stuff we need to make more rubbish. But who could have guessed, as the millennium trundles on to its close, that we would be running out of numbers? That was one resource everyone thought was infinite.

The numbers in short supply are telephone numbers. In some parts of the United States they are already quite scarce, and they will have to be carefully conserved over the next few years. At first the idea of such a shortage seems preposterous. A standard North American telephone number has ten digits: three for the area code, three for the central-office code and four for the local line number. A ten-digit format allows for ten billion distinguishable telephone numbers, from 000-000-0000 through 999-999-9999. Even if every person in North America had a telephone at home and at work, as well as separate numbers for a car phone, a fax machine, a modem and a beeper, there would still be more than enough numbers to go around.

The flaw in this analysis is that not all ten-digit numbers are possible telephone numbers. Indeed, more than 90 percent of them are unacceptable for one reason or another. A telephone number is not just an arbitrary sequence of digits, like the serial number on a ticket stub; it has a surprising amount of structure in it. As a matter of fact, the set of all valid North American telephone numbers constitutes a formal language, analogous to a computer programming language. When you dial a telephone, you are programming the largest machine on earth, the global telephone network.

A look at the grammatical structure of telephone numbers reveals a lot about how the telephone system works and how it evolved. And modifying that grammar turns out to be the key to solving the numbering crisis. The solution is discussed in a document released earlier this year by Bellcore, one of the surviving corporate fragments of the dismembered Bell System. The document has an imposing title: "North American Numbering Plan Administrator's Proposal on the Future of Numbering in World Zone 1."

When I was a boy, my grandmother's dialless telephone was an object of mystery. It was like a clock without hands or a ladder without rungs—I couldn't fathom the use of it. Then my grandmother demonstrated. She picked up the receiver and said, "Jenny, get me Mrs. Wilson, please. Thank you, dear."

My grandmother's telephone was already quite an anachronism when I first saw it in the 1950s. Automatic switching gear—allowing the customer to make a connection without the help of an operator—had been placed in service as early as 1892. The invention of the first telephone switch comes with a story worth telling. According to legend, Almon B. Strowger was a Kansas City undertaker who found he was losing business to a rival. Potential customers would telephone Strowger but "mistakenly" be connected to his competitor. Strowger noted that the competitor's wife was the switchboard operator for the local telephone system. His revenge was to invent a device that would eventually displace operators almost everywhere.

Strowger's invention was a ten-position rotary selector switch with a pivoting central arm that could rotate to connect with any of ten electrical contacts. The pivoting arm was moved by an arrangement of electromagnets, springs and ratchets. Each time the electromagnet received a pulse of current, it advanced the arm by one position. In the first network to try Strowger's idea, the customer operated the switch by pressing a button. If you wanted to dial a 7, you pressed a button seven times, thereby sending seven pulses of current to the electromagnet driving the selector arm. The push buttons were soon replaced by a rotary dial, which automated the counting of pulses.

A single switch of that kind could interconnect ten subscribers. If you were one of those subscribers, when you picked up your receiver, your line would be connected to the central selector arm. Dialing a one-digit number would then ring one of the other nine telephones. Adding a second stage of switching could expand the service to a hundred subscribers. Now the original switch, instead of being connected directly to ten subscriber lines, would be linked to a bank of ten more identical switches. Each subscriber would be identified by a two-digit telephone number. When you dialed the first digit, say a 3, the first selector switch would connect your line to the selector arm of the switch leading to lines 30 through 39. Dialing a second digit would move the selector arm of the second-stage switch to the appropriate contact.

It is easy to see how a Strowger switching network could be expanded to handle 1,000 lines (with three banks of switches) or 10,000 lines (with four banks). In principle, such growth could be continued indefinitely, but the quantity of switching gear would become impracticably large. The telephone company adopted a different plan. It set up switching offices that could each accommodate as many as 10,000 subscribers, then provided trunk lines to connect the various offices. At first, calls between central offices were completed by operators, but soon that task too was put in the hands of the customer.

Suppose you were a telephone subscriber when dialing between central offices was introduced. You were accustomed to ringing up your neighbors by dialing a four-digit number. Now you could reach towns and cities for miles around by dialing seven digits. The first three digits—actually two letters and a number at the time—specified the central office, and the last four digits were the line number within that office. But there was a price to pay: you would no longer be able to call next door by dialing just four digits. To reach local people you would have to dial your own central-office code before dialing the four-digit local number.

The extra dialing for local connections was a concern to telephone engineers, who worried that customers would resent it. Why couldn't you dial KLondike 5-2345 to reach your uncle across the river, but dial 5552 to reach your sister across the street? The question is a miniature version of a problem that has plagued telephone switching for at least forty years. In the first place, KL 5 is simply 555; the alphabetic encoding of numbers exists only on the dial of the telephone. Thus the first four digits of KL 5-2345 are the same as the local number 5552. When you have dialed those four digits, what should the switch do? Should it connect you to your sister, or should it wait to see if you dial more digits?

The Strowger switch allows little flexibility in resolving such ambiguities. It is called a step-by-step switch, because once it has made a selection, it cannot go back to revise the choice. A Strowger switch must determine on the basis of the first digit dialed whether to set up a local call or to select a trunk line for a call to another exchange. If the switch were to establish a tentative routing to your sister as you dialed 5552, there would be no way to undo that connection if you continued dialing.

Telephone switching gear has changed a great deal since Strowger's time. Modern switches are fully electronic rather than electromechanical, and they are capable of holding a series of digits in a buffer before determining what to do with them. Nevertheless, the architecture of telephone numbers is still strongly influenced by decisions made to accommodate the peculiarities of early step-by-step switches. Moreover, in some rural telephone office there may still be a Strowger switch clanking and clunking away.

Seven-digit dialing would seem, on first analysis, to give each telephone direct access to ten million others. Actually, the number of lines available is only about half that. The reason is that some numbers count for more than others.

"Dial 0 for Operator" has been standard telephone practice almost from the beginning of direct dialing. As a consequence, you will never see a telephone number such as 007-2345 or 099-6789, at least not in North America. If you tried to dial such a number, you would be connected to an operator before you could finish. That special handling of 0 puts off-limits a million potential phone numbers in every calling area—all the numbers from 000-0000 through 099-9999.

Another number you will never see for a North American telephone is 123-4567, in which the leading digit is 1. It turns out that various dialing codes beginning with 1 are reserved for internal uses within the telephone system, such as selecting trunk lines between switching centers. There go another million numbers.

The special functions of 0 and 1 forbid their use as the first digit of a central-office code, but traditionally 0 and 1 have been avoided as the second digit as well. At the outset that restriction had nothing to do with the operation of the switching network; instead it was a matter of mapping letters to numbers. Central-office codes were introduced with names rather than numbers because the telephone company thought BUtterfield 8 would be more memorable than 288. On the telephone dial, 0 and 1 are not assigned any alphabetic equivalents, and so they could not appear as the second letter of a central-office name. That subtle constraint, imposed to help avoid confusion between O and 0 and between I and 1, has had remarkably far-reaching consequences for the telephone system. Named exchanges are gone, but their influence on the format of telephone numbers remains.

For a long time 0 and 1 were avoided even as the third digit of central-office codes. There was no compelling reason for the practice, although again it helped avoid mistaking 0 for O or 1 for I.

In any case, for some decades most North American telephone numbers followed a pattern that can be expressed as *NNN-XXXX*, where *N* represents any of the eight digits from 2 through 9 and *X* is any decimal digit at all, from 0 through 9. The maximum capacity of this numbering system is equal to $8\times8\times8\times10\times10\times10\times10$, or 5,120,000. In practice, it is an upper limit that can be approached but not reached. A few lines in each central office are needed for testing and similar purposes, and a few exchange codes, such as 555, have traditionally been reserved. Moreover, telephone companies try never to fill a central office completely, since that would leave no flexibility when customers move or request a change in service.

By 1950 seven-digit dialing had spread to much of the U.S. (though not to my grandmother's house). A telephone connected to the network had the theoretical potential of reaching five million other phones. At the time there were fewer than fifty million telephones in the nation. Thus all that was needed, in order to allow a subscriber to reach out and touch everybody, was a factor-of-10 increase in the numbering capacity. One extra decimal digit would do it. The planners of the telephone system decided to be conservative. They came up with a scheme that would increase the capacity almost 150 times. A spokesman for one of the local Bell operating companies recently noted that under the plan the supply of numbers was expected to last for 300 years. It held out for almost fifty.

The idea, now familiar to all telephone users, was to divide the continent into area codes, known officially as numbering-plan area, or NPA, codes. In the original proposal, published in 1947, there were eighty-six assigned codes, with another fifty or so held in reserve for growth. Each state had at least one code to itself;

the more populous states had multiple codes. The largest cities were assigned codes such as 212, 312 and 213, which were the quickest to dial on a rotary-dial telephone. Every code had three digits.

To deflect resistance to the further lengthening of telephone numbers, the Bell System was careful to design the NPA codes so that extra digits would be needed only for dialing long distance; local calls could still be placed with seven digits, as in the past. Accordingly, the format of the NPA codes had to satisfy one fundamental requirement: the switching equipment had to be able to distinguish an NPA code from a central-office code. The key to making the distinction turned out to be the middle digit of the code. As noted above, the second digit of a central-office code had always been confined to the range 2 through 9; the corresponding digit of an NPA code is invariably either 0 or 1. Thus the middle digit alone distinguishes the two kinds of code. The complete pattern of an NPA code is *NZX*, where *N* again designates 2 through 9, *Z* is 0 or 1, and *X* is 0 through 9. The pattern for an entire telephone number is *NZX-NNN-XXXX*.

The *NZX* pattern yields 8×2×10 three-digit codes, for a total of 160. But as with the central-office codes, a few of the NPA codes were set aside for other purposes. Certain codes of the form *N*11 were reserved for reaching the telephone company itself (411 for directory assistance, 611 for repair, 811 for the business office); later 911 was added for emergency services. The numbers of the *N*00 series were designated service access codes instead of NPA codes; among various services offered, toll-free 800 calling has proved the most popular. Codes of the form *N*10 were given to the Telex network. Excluding all those twenty-four reserved codes left 136 combinations for ordinary geographic NPA codes. With 5.1 million numbers for each area code, the maximum capacity of the system was just under 700 million numbers.

Direct distance dialing with ten-digit numbers first went into service in 1951, in Englewood, New Jersey. If you were designing the switch for the central office in Englewood, how would you handle the challenge of the new extended numbers? Here is an informal description of one straightforward algorithm:

If the first digit is 0, connect to the operator.

If the first digit is 1, signal an error.

Otherwise, since the first digit is in the range 2 through 9, examine the second digit. If it is anything other than 0 or 1, the dialed number must be a local one, and so it can be handled by the established seven-digit protocol.

If the second digit is 0, examine the third digit. If the third digit is also 0 (so that the first three digits form the pattern *N*00), connect to the appropriate special service. Otherwise (the pattern being *N*01 or *N*0*N*), connect over the long-distance network to the appropriate NPA.

If the second digit is 1, examine the third digit. If the third digit is 0 or 1, connect to the indicated *N*11 or *N*10 service. Otherwise (the pattern being *N*1*N*), connect over the long-distance network to the appropriate NPA.

An algorithm like this one worked in Englewood, but a problem showed up when telephone engineers elsewhere tried to implement it. The algorithm will not work on a step-by-step switch. As I noted above, such a switch must commit itself to the routing of a call when the first digit is dialed, but the algorithm offered here cannot know whether a number is local or long distance until the second digit has been examined. Deferring the decision was not a problem in Englewood, because the central office there had a switch capable of storing the first few dialed digits in a buffer, but many other switching offices still relied on step-by-step equipment.

A couple of solutions to the step-by-step problem were tried. The scheme that caught on was to ask the customer to dial yet another digit for long-distance service, namely a 1 before the area code, a practice that came to be known as 1+ dialing. (The plus sign is not dialed, of course; it is meant to suggest that the 1 is only a prefix code.) Since 1 could not be the initial digit of any valid NPA code or central-office code, the new marker was unambiguous.

There is something ironic about the introduction of 1+ dialing. The planners of the telephone network had taken pains to design area codes that could be distinguished automatically from central-office codes, but 1+ dialing made the internal distinction redundant. If the 1+ prefix had been part of the plan from the outset, there would have been no need to restrict NPA codes to the *NZX* format, and the codes could have been supplied more liberally. For a time, a faction within the Bell System hoped and expected that 1+ dialing would eventually disappear. In their view the NPA coding was an elegant and parsimonious scheme that cleverly exploited all the peculiarities of the existing switching network to extract the maximum amount of information from the minimum number of digits. In contrast, 1+ dialing was a crude and wasteful patch that should be dispensed with as soon as the last step-by-step switching plant was scrapped. But the patch is still with us, and it has patches of its own now.

Number shortages are nothing new in the larger metropolitan areas. New York, Los Angeles and Chicago have been struggling for years to eke out the supply. One of the first steps taken when an area code starts to fill up is to expand the list of central-office codes to include numbers of the format *NNX*, rather than just the ones that conform to the template *NNN*. In other words, exchanges ending in 0 and 1 are allowed. The change is painless, since the third digit of a central-office code carries no special significance anyway. It increases the number of available codes from 512 to 640, a gain of 25 percent.

The next recourse is to allow central-office codes of the form *NXX*, where both the second and the third digits can be any number, including 0 or 1. That easing of the rules creates 160 more codes, bringing the total to 800, but it also has a grave consequence. It eliminates the structural distinction between central-office codes and NPA codes. Once a network has introduced *NXX* office codes, some kind of extra signal, such as 1+ dialing, is all but mandatory to distinguish local from long-distance calls. There is no going back to the original plan of eliminating ambiguity by examining the second digit of the dialed sequence. Los Angeles was the first city to adopt *NXX* central-office codes, in 1973. New York held out until 1980, then Chicago followed in 1983. By now about twenty areas have converted to *NXX*.

NXX exchanges yield eight million subscriber numbers for each area code. When that supply proves insufficient, the only option is to split the area and introduce a new NPA code. That process began soon after direct distance dialing was launched, and by the late 1980s it had become apparent that all 136 of the available NPA codes would soon be allocated. Growth in demand was not abating. Where could more numbers be found? A stopgap was to recover some of the *N*10 codes that had been assigned to the Telex network. They all were returned except 610, which is still used by the Canadian Telex system, and 710, whose function is now listed as Government Special Services. NPA codes 310, 410 and 510 are already in service, and they will soon be joined by 210 and 810. At that point World Zone 1 will have only one NPA code left: 910.

The United Nations agency that regulates international telecommunications divides the world into nine zones. World Zone 1 includes the U.S. and Canada and about a dozen Caribbean nations. There are eight other world zones: 2 is Africa; 3 and 4 cover Europe; 5 is Central and South America; 6 is the Pacific; 7 is the territory of the former U.S.S.R.; 8 is Asia; and 9 is the Middle East.

Within Zone 1 the administration of numbering has been delegated to Bellcore, which was therefore expected to find a solution to the NPA code shortage. The heart of Bellcore's plan is to relax the syntactic constraints on the form of an

NPA code. Specifically, area codes, like the newest central-office codes, are to have the format *NXX*: the middle digit can be any number, not just 0 or 1. That change yields a fivefold increase in the number of possible codes, from 160 to 800. Of the 640 new codes, Bellcore proposes that 300 be held for use as ordinary geographic codes. Thus the capacity of the system would triple. With an eventual total of 442 area codes, each using *NXX* office codes, there would be room for 3.5 billion telephone numbers.

Another ninety of the new codes are earmarked for nongeographic services, such as the existing *N*00 series of service access codes. That large allocation reflects the tremendous success of 800 service (AT&T recently reported that 40 percent of its long-distance calls go to 800 numbers) and the more recent popularity of 900 service. Perhaps even more important is the advent of "personal communications numbers," or numbers associated with a person rather than a telephone. AT&T recently introduced service of that kind—a number that follows you wherever you go—keyed to the 700 service access code. Such applications can have a double impact on the demand for telephone numbers. The way the telephone system now works, when you call an 800 number, that number is looked up in a data base, which records the "real" telephone number with which each 800 number is associated; then your call is passed along to the second number. Thus for every 800 number there is at least one ordinary number needed as well. Unless that arrangement changes, filling up ninety service access codes will also fill up another ninety ordinary NPA codes.

After those allocations for geographic and nongeographic codes, 250 numbers remain. Bellcore recommends that 170 of them be set aside for events and needs that simply cannot be foreseen. The last eighty codes would be held in reserve, to be applied when even the expanded supply of numbers is finally exhausted. At that point there will be no choice but to add more digits to phone numbers.

The adoption of *NXX*-format area codes will eliminate all distinctions between an area code and a central-office code. How will switches tell them apart? One possibility is to continue requiring a 1+ prefix on any ten-digit call but to forbid 1+ on all seven-digit calls. The Bellcore plan recommends a different approach: it would require a ten-digit number for every call, including local calls. Then the switch could always treat the first three digits dialed as an area code, the next three digits as a central-office code and the final four digits as a customer line. The 1+ prefix could be dropped, since there would be no need to alert the switch that ten digits are coming.

The Bellcore plan is a thoughtful and circumspect document, which carefully acknowledges all the hazards and limitations of technological forecasting. It was prepared with the advice of some forty "experts and futurists," and it doubtless also draws on quantitative analyses of population growth and of trends in the telecommunications industry. Still, I cannot help wondering if it might not represent another major miscalculation.

The planners of the 1940s underestimated the demand for telephone numbers because they could not foresee the variety of ways those numbers would be used. At that time a telephone was a black box permanently wired to the wall, and nothing other than a telephone was ever plugged in to the Bell network. No one anticipated the proliferation of modems and fax machines—or of telephones that don't plug in at all. The Bell System engineers never dreamed that people would chat on the phone while strolling through the supermarket or tanning on the beach or plowing a cornfield. They never guessed that paging devices (each with its own phone number) would be carried not only by doctors on call but also by plumbers and professors and street-corner cocaine dealers.

Claes Oldenburg, Soft Pay-Telephone, *1963*

The mistake that now seems hard to avoid is assuming that the demand for telephone services, and particularly for numbers, will continue to grow in the same way. I have no doubt that communications traffic of all kinds will increase dramatically. But it seems possible that some substantial fraction of the traffic will be diverted from the telephone system into other channels. The coming decades will surely bring communications devices just as unexpected as the fax machine, the cellular telephone and the beeper, but it should not be taken for granted that those devices all will have telephone numbers.

The telephone system is a circuit-switched network. For most of the history of the system, when you placed a call, you were renting a pair of copper wires that ran continuously from your telephone to the other party's phone. You had exclusive use of those wires during the call; when you hung up, they were rented to someone else. Today the transaction is more complicated (your call may well share a fiber-optic cable or a satellite with hundreds of other calls), but conceptually the system still works the same way. When you dial the phone, you get a private connection to one other party.

There is an alternative network architecture called packet switching, in which all stations are always connected to the network, but they receive only the messages addressed to them. It is as if your telephone were always tuned in to thousands of conversations going by on the wire, but you heard only the occasional word intended for you. Most computer networks employ packet switching, because it is more efficient than circuit switching when traffic is heavy. It seems reasonable that the existing packet-switched networks will grow, and new ones may be created; they could well absorb traffic that would otherwise go to the telephone system, and thereby reduce the demand for telephone numbers.

As the architecture of communications networks changes, so will the user interface. Telephone numbers may eventually become obscure internal codes that the general public has no need to know. Already many telephones come with speed-dialing buttons so that you can record frequently called numbers (and thereafter forget them). There are also pocket-size dialers you hold up to the mouthpiece of a telephone. If you wish, the telephone company will store your list of favorite numbers, so that you can dial them with a one- or two-digit code. Such strategies for insulating the customer from the number itself will become more prevalent as numbers grow longer and harder to remember. I can imagine a kind of user interface that might ultimately evolve. In a couple of decades, perhaps, the telephone will have no dial at all. You will simply pick up the receiver and say, "Jenny, get me Mrs. Wilson, please. Thank you, dear." ●

BRIAN HAYES is editor-at-large of AMERICAN SCIENTIST.