CHAPTER 1

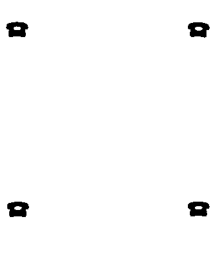
Ronald L. Graham

SOME YEARS AGO the Bell Telephone Company had, in a room in one of its New York offices, a very large map of the United States. This map showed all of the major cities with long-distance service, connected by strings representing telephone lines. When one calls long-distance between New York and Los Angeles, for example, the call may be routed through Atlanta or Detroit—or through many other cities, depending on the availability of circuits. These routes can differ in length by hundreds, if not thousands, of miles, but one is charged the same amount—based on the minimum distance—no matter what path the call actually takes. To determine the shortest route of a telephone call, if it went over existing facilities, the lengths of these strings were measured and compared to other possible routes.

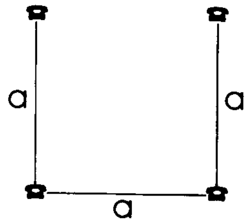
By the mid-1950's, after Bell Labs had taken delivery of the largest computers being manufactured by I.B.M., the question arose as to whether the strings could be replaced by some model calculation done with a computer. Such a calculational procedure is called an "algorithm," and, in 1957, the Bell mathematician Robert C. Prim, building on work done a year earlier at Bell Labs by Joseph B. Kruskal, found a simple algorithm that was well-suited to machine computation. The strings have now disappeared.

Roughly speaking, there are three kinds of computational problems. In 1936 the British mathematician Alan Turing demonstrated the existence of problems for which no algorithmic solutions exist in principle. They are closely related to the Gödel theorems, which demonstrate the undecidability of some propositions in formal mathematics. Then there are problems for which an algorithm exists—the problem is "decidable"—but the computer time needed to use it would grow exponentially with the "size" of the problem. For all practical purposes such an algorithm is useless, since it might take, for example, a time equivalent to the present age of the universe (some 12 billion years) for a computer—or, for that matter, any realistic number of computers—to produce a solution. Such problems are called "intractable." Finally, there is a class of problems—mathematicians working in this field tend to call them the "easy problems"—for which there exists an algorithm of such a character that the computer time does not grow faster than some polynomial in the size of the problem. For example, if the size is n, then the computer time might grow only as fast as n^2.

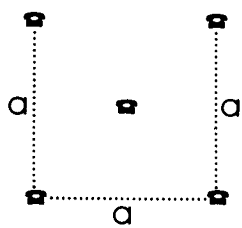
Although all of this might seem somewhat academic as far as the telephone company is concerned, it is not. The fixing of tariffs for "Private Line" telephone service is a practical application of such algorithms. Many large corporations have private lines connecting their operations in various locations. The history of how tariffs for these lines were arrived at seems to be rather obscure, but one factor has to do with minimizing the length of the telephone lines connecting the locations. Below is a diagram of four locations, which, for simplicity, I have put on the corners of a square.



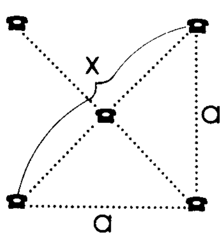
Let us assume that existing telephone lines connect these points. As the next diagram shows, at least three lines are needed in order to connect all four points. If the edge of the square has length a, then the total length needed to connect these points is $3a$.



There is nothing very obscure about this, but this is not how the effective length a was computed for determining the tariff. To illustrate how this was done, I have put a hypothetical point in the middle of the square. This point need not correspond to an actual point on the telephone network; in fact, it generally won't correspond to such a point. There need not even be a telephone there. It is simply an arbitrarily chosen point that could, in principle, correspond to a telephone exchange.

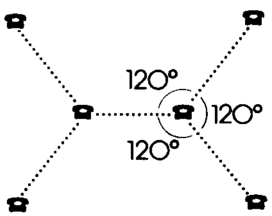


In the next diagram I have connected the imaginary telephone exchange by four imaginary telephone lines to the corners of the square.



By using Pythagoras' theorem, one can readily show that the length of each of these imaginary lines is $\frac{1}{\sqrt{2}}a$, and since there are four of them the length of the imaginary network is $2.8a$, which is less than

$3a$ by about 7 percent. It turns out that one can do even better by adding *two* imaginary points, as shown in the next diagram (the central lines meet at 120°):



The length of this imaginary network is $(\sqrt{3} + 1)a$, or $2.7a$. This, it can be shown, is the *best* that one can do, no matter how many imaginary points are added to the square, and according to the tariff structure, this "hypothetical length" is used to figure out the tariff. AT&T was forced to do this because of successful lawsuits that argued that Private Line tariffs should be based on the minimal-possible-length telephone networks and not on some network chosen for the convenience of the telephone company.

Such a network is called a "tree" by mathematicians, and these imaginary extra points are named "Steiner points" after the nineteenth-century Swiss mathematician Jakob Steiner, who studied this problem for the case of three points. The tree that, in the case that we have been considering, produced the smallest net lengths is called the "Steiner minimal tree." The last diagram is the Steiner minimal tree for this case. For certain very simple, symmetrical cases, like the one here, determining the Steiner minimal tree is an "easy problem." But if more cities are added to the Private Line network, and if they are not located symmetrically, determining the minimal Steiner tree exactly is, in a practical sense, hopeless. As Ronald Graham, the Bell mathematician who first introduced me to this problem, put it, "You could have a New Jersey full of Crays [the fastest computer now operating], and you still couldn't solve the general Steiner problem for twenty-five points." Thus, there was no rigorous way in which AT&T could have decided precisely what tariff to charge for Private Line service where a great number of locations was involved.

It has never been clear to me what is meant when someone says so and so *looks* like a mathematician. However, I am quite sure that Ronald Graham does *not* look like a mathematician. He looks, perhaps, like a professional acrobat or juggler, both of which he was. He was once president of the International Jugglers Association, and he is the only mathematician I know of who worked his way through graduate school, Berkeley in this case, by performing in a circus with a trampoline troupe. For some time Graham had a large net suspended from his office ceiling, to catch stray objects dropped while he practiced juggling. (Claude Shannon, the former Bell electrical engineer and computer scientist, who was one of the creators of information theory, was able to ride a unicycle and juggle simultaneously.) Graham, who is blond and about six feet two, and obviously in superb physical condition—a rarity among mathematicians—has recently taken up tennis and immediately became a favorite to win the Bell Labs tennis championship. He is also by all accounts a superb mathematician. He was born on 31 October 1935, in Taft, California, where his father worked in the oil fields. He has been at the Laboratories since 1962 and is now director of the Mathematics and Statistics Center, a group of about sixty mathematicians some of whom work, as does Graham, in discrete mathematics—"discrete" as opposed to "continuous"—which includes the kind of graph-theory problem represented in finding the Steiner minimal tree. The Center publishes something like 250 papers a year, and Graham, alone or with collaborators, publishes about fifteen. In addition, he edits twenty-five journals in computer science, combinatorics, number theory, operations research, graph theory, and general mathematics. He also holds a rather bizarre record, which appears in *The Guinness Book of World Records,* for the largest number ever used in a mathematical proof. As Graham remarked, "It is too large to understand what it actually means." He came across it in 1977. It is inexpressible without special notation and is, appropriately, known as Graham's number.

During Graham's childhood, his peripatetic father moved back and forth between California and Georgia, changing jobs frequently. He finally joined the merchant marine, and he and Graham's mother were divorced. She moved to Florida, where Graham was put into yet another school. But he was a very gifted student and left high school after the eleventh grade with a Ford Foundation fellowship, which he used to enter the University of Chicago at the age of fifteen. In the early 1950's Chicago was involved in one of those educational experiments in which students are taught physics by reading the original works of Isaac Newton—generally a losing move. Although Graham was nominally a science major, he did not take a single course in mathematics. After three years his father offered him the chance to go to a somewhat less adventurous college if he would move back to California. So Graham transferred to the University of California at Berkeley, where he majored in electrical engineering. After a year there, he still had no degree, although he had done the research for his Ph.D. thesis. However, he was eligible for the draft, so he enlisted in the air force and was sent to Alaska. During his off-hours, he attended the University of Alaska in Fairbanks, where he took his degree in physics in 1958—seven years after entering the University of Chicago. After leaving the air force, he now had the means to return to Berkeley and begin his formal graduate work, although he had essentially done the research for his Ph.D. thesis.

Berkeley has a rather traditional mathematics department. As Graham put it, "The department is quite pure. There is a strong emphasis for people who graduate from there to teach and that is what I thought I would be doing; but I wanted to take a look at the other side too." The "other side," in this case, was mathematics as done in an industrial laboratory. In 1962, when Graham got his degree, David Slepian, a Bell Labs mathematician, was traveling around to the major universities to recruit mathematicians for the Laboratories. As Graham explained, "Bell Labs has a whole army of people in different disciplines who are assigned to go to specific schools and talk to potential employees. But I was told by the people at Berkeley that if I went nonacademic I would be mathematically dead in three years. I decided to come anyway, with the idea of trying it out for a few years, making a little money and then deciding what I'd *really* do. I'm still here—twenty-two years later. Slepian was a pretty persuasive guy."

I was curious to know whether Graham was recruited for any particular project at Bell. Mathematicians are so idiosyncratic that it was hard for me to imagine them being recruited to work in a group on some specific project. Graham explained that at the Labs about 10 percent of the work in general is officially designated as "research," the rest being in developmental areas. The Mathematics Center, by and large, does only undirected research. As Graham put it, "Our basic philosophy is to get the best people we can and, in some sense, to stay out of their way. I think in my case Slepian hoped I would get involved in what is called 'encoding theory'—a branch of mathematics of which, along with Shannon and some of the other people here, he had been one of the inventors. It involved techniques that were similar to what I had been using for my thesis. Slepian gave me a book to read and suggested that I might lead a seminar on the book. I looked through it and noted very quickly that many of the people who had created the field were right here, so I thought that my giving a seminar on the subject was, to put it mildly, a little redundant. I learned the material, which was very useful to me later for other work, but we ended up doing a seminar on something else."

Graham, who has had the responsibility for bringing people into his group, told me, "Typically, it takes someone one or two years to learn the ropes well enough here, to get a feel for what goes on and how to function in this environment. It takes time for people to get plugged into the various networks around here. One might mention to a new person a few of the problems that are floating around. One difference between Bell Labs and a university is that here, more or less, everyone comes in every day and the office doors are open. There is a lot more interaction here than in most universities, and a lot less feeling that the problem I am working on is '*my*' problem and that I'll tell you about it when I've dotted all the *i*'s and crossed all the *t*'s. We have much more of a community effort here, and it crosses disciplines. For example, some of the chemists are now looking at the structure of graphs—the kind of thing I do—and they like to get hold of mathematicians to try out their ideas.

"But the atmosphere you work in here is very much a local phenomenon. It depends in a fairly strong way on your local supervision. For example, there are certain regulations about how much vacation you get each year, and there are even regulations about your daily starting time. There are some division managers who apply these regulations pretty literally. Three weeks' vacation doesn't mean four. But within the Mathematics Center it has been a tradition to let each person function the way they do best. We don't expect someone to come in each day and prove two theorems and a corollary before lunch. If you're in the mood, and things are rolling, and you think you have the right insight, you can work. If a problem has really gotten to you, you know you are going to work on it night and day. If not, you are probably going to do a certain amount of paper-shuffling to pass time. Of course in universities you have teaching, which, in a sense, can justify why you are getting paid even when your research is not going well."

This discussion raised a basic question about the Laboratories, and I wanted to get Graham's reaction to it. Universities have a built-in mechanism for dealing with people who have passed their peak research years: they can do more teaching and administration. I was also aware that mathematicians, like many scientists working in these more abstract fields, tend to "burn out" after their forties. Indeed, one of the great problems confronting all but the very best universities is that the average age of their department members is steadily increasing, owing to a mixture of tenure and economics. Tenure guarantees a job for life, and there is not enough additional money to hire young people. Hence many departments are more and more filled up with aging, tenured professors who are, at least in these very abstract fields, less and less productive. Graham said that the average age at the Mathematics Center was definitely below forty. This raised the obvious question of how this favorable age distribution was maintained. He responded, "There is one fundamental way in which Bell Labs is different from a university. There is no tenure here. There is what you might call a weak moral tenure. However, there certainly have been people here who have done good work for a number of years early in their careers, but as time went on they did less and less, until finally, so to speak, they just somehow died. Such people need some real new impetus to get them going again, and there is certainly the mechanism here to provide that. Our vice-president, for example, could say, 'Now you are over *there* in, for example, a developmental area,' which means you are either 'over *there*' or you are out. But I think that the strong factor that motivates people here comes from within the people themselves. Good people here aren't held with their noses to the grindstone.

"No one says, 'You will now do this problem, and then you will do that problem.' Because if you want the very best people to do their best work you must give them an environment that allows them to do that. I think what happens here is that after a few years, during which you have been doing work and getting a reasonable salary with no onerous demands placed on you—time and energy—you can feel something that I wouldn't say is quite guilt. It's rather more like an internal motivation. You say to yourself, 'Let me take another look to see if there is something I can do now that would be really useful for the good of Bell Laboratories.' That's actually quite a strong motivating factor. Many people here go through this kind of soul-searching during their first few years here.

"They ask themselves, 'Why am I here? What am I doing here? Why am I being paid?' and so on. The Mathematics Center has traditionally been a breeding ground for upper-management personnel. People often get promoted kind of diagonally—up and out of pure research—or get put into other areas that are more development-oriented, which don't make the same kinds of demands on one's research abilities. This has certainly been responsible for some of the bias we have in our age distribution in the center."

Graham's own research spans both pure and applied mathematics. In 1972 he shared with two other mathematicians the prestigious Polya Prize for his work in "Ramsey theory," an odd branch of pure mathematics that has to do with finding unexpected order in apparently random mathematical situations. For example, if one arranges the numbers 1 through 101 in any random order, the theory guarantees that there will always be at least eleven numbers arranged in increasing order or at least eleven in decreasing order, so, to that extent, no arrangement is entirely random. Until recently no one had found any application for results like this; but Ramsey theory is now being used in the design of data networks. These are, by and large, the sort of intellectual curiosities that mathematicians delight in.

In contrast, much of Graham's more recent work has had to do with the kind of practical problems that arise from situations like setting Private Line tariffs. Indeed, he initiated a branch of mathematics that he calls the "worst-case analysis." He became interested in this in the 1960's, when Bell Laboratories got involved in writing the software for the then-proposed antiballistic-missile system. Graham knew some of the people working on the problem, which basically involved how to optimize the scheduling of a large number of interrelated tasks—in this case, identifying and locating fleets of incoming missiles. They had discovered that the order in which these tasks were performed made a crucial difference to the end result. More surprising, they also had discovered that an order that appeared good could really be very bad, if each task took a certain amount of time, and if this time were reduced. Prior to this discovery, it was assumed that by adding computers, thereby reducing the waiting time for certain jobs to get started, it would be possible to reduce the time it took to do the whole job. It turned out that adding more computers could actually make things worse. There might be no limit to how bad you could make the overall efficiency simply by doing what, at first, seemed reasonable: bringing more computers to bear on the problem. Graham was able to prove that this was not so. There is a limit to how badly one can make the overall efficiency by adding computers in a problem like this. The fact that such a limit exists enabled Graham to make what he calls "performance guarantees." If one follows certain rules, no matter how quickly or in what order various tasks are done, the end result cannot depart from the best way of doing the tasks by a certain percentage, which, in many cases, Graham could calculate explicitly. One of the situations he analyzed was the NASA Apollo Mission, where three astronauts were asked to accomplish various tasks. The question arose how much worse would their overall performance have been if, following a certain set of rules, they had carried out their tasks in the worst possible way. Among the rules was one that said that if an astronaut were free to do something he should be doing it and once started on that task he should finish it. In this case Graham showed that the worst case was only 40 percent less efficient than the best case.

There is another example of this kind of analysis that Graham likes to give because it is simple to state although very nontrivial to study. Suppose that one is given five weights, two of them weighing three pounds and three of them weighing two pounds. Is it possible to find a strategy—an algorithm—that can be used to divide the weights into two piles which have nearly the same weight. This case is so simple that by trial and error—which is hardly an algorithm—one sees that the solution is to have a pile with two weights weighing three pounds and another pile with three weights weighing two pounds. Each pile will then have a total weight of six pounds. But suppose one wishes to give a systematic method for doing this and one insists that this method will give the best answer in the general case involving any collection of weights. As far as anyone knows this is not a "tractable" problem. There is no known algorithm that would enable a computer, or, as Graham put it, even "a New Jersey full of Crays" to solve the general case in any reasonable time. There are algorithms for sorting them into piles, but there is no guarantee that any of these algorithms will produce the best solution. Graham gave me an example of a simple algorithm, which can be applied to the five weights discussed above.

1. First take the weights and arrange them in order of decreasing weight. We would have then the five weights arranged in the order 3,3,2,2,2.

2. Now we begin arranging these weights into two piles by the strategy of putting the heaviest weight in one pile and the second heaviest in the other pile. So in this case we have one 3-pound weight in each of the two piles.

3. Take the next heaviest weight, 2, and put it into the lighter pile which will tend to even things out. We now have the piles arranged 3,2 and 3,2.

4. Keep repeating the last instruction until all the weights have been placed. In our case we end up with the piles 3,2,2 and 3,2. But this has produced an uneven division, and we know that the best solution is an even division 3,3 and 2,2,2. How much of a mistake have we made? The algorithm gives us a heavy pile of 7 while the "correct" answer—the best answer—is 6 so we are 7/6 off. We can see all of this by inspection here, but Graham proved that if one used this algorithm for *any* collection of weights one would never do worse than being off by this same factor of 7/6. The algorithm will not be wrong by more than about 14 percent which is the "performance guarantee" for this algorithm.

Like so many of the people I talked to at Bell Labs, Graham is worried that the divestiture will change the character of the place—and not for the better. The hiring pattern at the Mathematics Center has already changed: there is now an almost exclusive emphasis on experts in computer science, a very broad subject covering everything from artificial intelligence to the kind of scheduling problems on which Graham is such an expert. Indeed, shortly after my visit he was on his way to the California Institute of Technology, where he functioned as a kind of "interface" between computer scientists and pure mathematicians. He also gave some of his informal "seminars" in juggling, which are immensely popular with students.

Graham told me that the universities are now producing something like 200 Ph.D.'s a year in computer science, well below what companies like AT&T and IBM—let alone, all the universities—can absorb. He worries that industry will gobble up the entire crop of computer science Ph.D.'s, leaving no one to teach. Few universities can compete with the salaries and working conditions that a place like Bell Labs can offer. At Graham's suggestion, some members of his own group have been split off into a computer science group within the Mathematics and Statistics Center—in the hope of strengthening ties between mathematics and theoretical computer science—and Bell Labs has just created a house organ in which they can publish. The *Bell System Technical Journal* may be unique among industrial publications, in that it is produced by a company but contains results of such fundamental importance that workers in certain fields must read it regularly if they are to keep up. At Stanford University, where Graham taught in 1982, the mathematics department had only five or six new graduate students in mathematics; computer science, a very strong specialty at Stanford, had hundreds of applicants. "People know where the exciting problems are," he remarked, "and they have to eat too."

Graham gave me an example of the kind of thing that worries him about the future at Bell Labs. A member of his group is one of the world's greatest experts in the theory of probability. Some years ago his son was stricken with a brain tumor. Graham's colleague became intensely interested in how such tumors can be detected with CAT scanners. This is basically a mathematics problem—the reconstruction of a three-dimensional volume from two-dimensional projections of it. The scanner produces various projections of a mathematical function and the problem is to find an algorithm by which the function itself can be reconstructed from these projections. This is the practical problem of using such a scanner to locate a diseased area in a three dimensional volume. Bell Laboratories is not in the tomography business, although it does do research on machines used for medical diagnostics. (For example, the standard machine used for the bilirubin test on blood plasma—to test liver function—was invented at Bell.) At first, some of the upper management was not very pleased with the proposed switchover, but Graham's colleague persisted, and on his own initiative he spent six months at a Boston hospital studying tomography. Ultimately he became one of the world's leading experts in computer tomography and recently won an important prize for this work. Graham said this was then "not an atypical sort of switchover of fields for someone who, as in this case, was extremely talented and whom you really trusted. Our feeling was that what he was going to do might take a few years, but that was O.K. We would still be here. Now there is a certain concern, hard to put your finger on exactly, that this long-view atmosphere may get shortened. When I came here there was the attitude that although you were not supposed to do things that were completely irrelevant to the basic mission of the Labs—communication—it was all right if you did something that might have a feedback in ten, fifteen, or even twenty years. We have many examples of things that someone did, and it was only twenty years later that we found that we really needed it, that it was a good thing we had it. The window may now shrink to five years, three years . . . next month.

"There is a term that is used around Bell Laboratories—'fire fighting'—which is what a lot of the people at Murray Hill and other places such as Holmdel do a fair amount of. They have specific problems with deadlines. For example, just down the hall there are some people working on an electron-beam method that will be used to etch circuits on chips. They have a very precise deadline. But that kind of research has limitations. The major developments are unexpected. If you really knew what you were trying to do, that would often be the biggest part of the battle. There does not seem to be any obvious way of knowing how some development here will impact on something over there. You just hope you have good people who are excited and that they can communicate."