THE GLOBAL POSITIONING SYSTEM

Two dozen satellites hovering thousands of miles out in space are allowing people to locate themselves on the earth's surface with remarkable precision

Turning onto the final approach the Boeing 737 airliner responded smoothly to the command of its computerized autopilot, setting up for what looked to be another perfect landing. Although automatic approaches are routinely performed in bad weather, this particular aircraft was not using the normal navigational signals beamed up from the airport to complete its so-called Category IIIA landing--the kind used when the pilot cannot see the runway until after the airplane touches down. The jet's occupants were relying instead on satellites of the U.S. Department of Defense's Global Positioning System (GPS) high in orbit overhead. These modern navigational benchmarks, floating in space at an altitude of more than 20,000 kilometers, were supposed to guide the swiftly moving aircraft safely to the ground.

As the 737 neared the runway, the GPS signals indicated that the ground loomed only 300 feet below the landing gear, and the airliner slowed its descent. Having completed numerous landings that day, the engineers on board had grown confident in the craft's satellite-guided abilities. But on this attempt the autopilot suddenly sounded an alarm: the GPS equipment had lost contact with a critical satellite. The airplane's human pilot quickly took over control from the computer and throttled up the engines to abort a potentially disastrous landing.

After later analysis, the engineers who had developed the airplane's guidance equipment realized that the temporary loss of signal had been caused by a software "bug" in the GPS satellite. The flaw had not been detected earlier, because no one before had relied on GPS navigation for such a demanding task. The GPS system had, in fact, been designed with a built-in uncertainty in position of 100 meters--longer than a football field. Several years ago few people would have dared to imagine that the GPS could lead an airplane all the way to the ground. But in the intervening time some clever tinkering has brought about a surprising level of navigational precision, and the GPS has evolved into something even its makers had not envisioned when they launched the first of its satellites.
The Department of Defense began construction of its sophisticated satellite positioning system in the mid-1970s to allow military ships, aircraft and ground vehicles to determine a location anywhere in the world. Although the designers of the GPS had meant it primarily for classified operation, they made provision for civilians to use the satellite signals to locate themselves—but far less well than their military counterparts. A reduction in accuracy seemed necessary for the unclassified signals; otherwise an enemy could easily use the GPS broadcasts, and the elaborate satellite system would not give the U.S. any military advantage. Yet remarkably, scientists and engineers working outside the armed forces have devised ways to circumvent the purposeful degradation of the GPS signals, and ordinary citizens are now able to achieve much better results than the Department of Defense had ever expected.

Such refinements allow GPS radio receivers to guide pleasure boats into foggy harbors or passenger cars along unfamiliar roads. This satellite positioning system is now even used to keep track of the placement of cargo containers as they are shuffled around the holding yards of Singapore’s busy port. More impressively, with GPS surveying equipment, geologists can measure the subtle shifts in the earth's crust—movements of just a few millimeters—that show the motion of the planet's tectonic plates and help to define the location and extent of earthquake-prone zones.

The glitches that sometimes emerge during such exercises are not so much failures as they are indications that nonmilitary Scientists and engineers are pushing GPS instrumentation to limits never intended by the system's originators. How did the Department of Defense intend to deny their signal's inherent precision to civilians? Why have so many people succeeded in circumventing the prohibitions on accuracy? The answers to these questions require a broad understanding of the mechanics of satellite navigation in general and of the GPS in particular.

**A New Star to Steer By**

Soon after the Soviets launched Sputnik in 1957, some scientists and engineers realized that radio transmissions from a satellite in a well-defined orbit could indicate the positron of a receiver on the ground. The procedure uses the Doppler shift of radio signals as the satellite passes overhead. (A similar Doppler shift accounts for the sudden change in the tone of a train whistle as a locomotive speeds by.) Using this method, the U.S. Navy pioneered the "Transit" satellite positioning system during the 1960s.

The technique the navy employed was, unfortunately, rather cumbersome. It required expensive electronic equipment on the ground and usually demanded reception of signals from two separate passes of the satellite overhead, which necessitated a wait of more than 100 minutes. Even under the best circumstances, with several days available for collecting signals, one could not hope to determine a location that was more accurate than about a single meter, and so Doppler positioning proved rather limited for precise land surveying.

But even before the deployment of the first Transit satellite, the Department of Defense had begun contemplating a more sophisticated approach that might, for example, allow a pilot flying a jet fighter instantaneously to determine his exact position. In particular, the U.S. Air Force was planning a navigation system that utilized "ranging"—the measurements of distances to several satellites—rather than the Doppler shift in radio frequency.

The determination of positron by ranging is straightforward in concept. Suppose, for example, one is able to ascertain that a particular satellite is 20,000 kilometers away. Then the person's position must be somewhere on a huge sphere 20,000 kilometers in radius (40,000 kilometers in diameter) that surrounds that satellite. Because satellites travel in stable, predictable orbits, the location of the satellite, and the imaginary sphere surrounding it, is known exactly.

If at the same instant that the first range is taken the person can also measure the distance to a second satellite, a second "sphere of positron" can be determined. A third range to a third satellite gives a third sphere, and so forth. In general, there will be few places where all the spheres meet. For example, two spheres can intersect along a circle; three spheres can coincide only at two points. Because one of these points typically represents an unreasonable solution to the navigation problem (it may be deep within the earth or far out in space), three satellite ranges are sufficient to give one's exact positron.

**Synchronize Your Watches**
The question the military planners first faced in designing the GPS satellite positioning system was how exactly to make the necessary range measurements. At that early juncture, there were many choices. For example, radar equipment could transmit a radio pulse and receive the echo after the signal had propagated up to a satellite and reflected back down again. A computer could then easily calculate the distance to the satellite from the measured delay and the known velocity of the radio pulse, the speed of light.

But such a system would force anyone using it to broadcast a stream of powerful radar bursts—not an ideal activity for soldiers, sailors or pilots who are trying to avoid being detected by their enemies. So the Department of Defense considered an alternative strategy. The navigation satellites could transmit radio pulses at specific, known times, and by measuring the exact instant when the pulses arrived, the receiving equipment could determine the distance to the satellite. That procedure demanded, however, that the receiver's clock be synchronized with the one on the satellite. This concept formed the basis for what became the Global Positioning System.

Exact synchronization may at first seem a rather severe requirement; a mismatch of as little as a millionth of a second would translate to an error of about 300 meters. Although the navigation satellites themselves could each carry a highly accurate "atomic clock," it would be prohibitively complicated and expensive for each receiver to be so equipped. But there was a way to avoid the need for such perfect timepieces on the ground: one need only establish how much the receiver's inferior clock had drifted from the correct time.

This task is not particularly difficult. One starts by assuming that the receiver's clock is approximately correct in calculating the ranges to four satellites. Because the receiver's clock is not in fact running exactly on time, the distances calculated, called pseudo-ranges, will not be entirely correct. The four pseudo-ranges will correspond to four imaginary spheres surrounding the satellites. These four spheres should ideally intersect at a single point—the receiver's location—but will not meet exactly, because the satellite and receiver clocks are not absolutely synchronized. All four spheres will be just a little too large (if the receiver's clock is running fast) or too small (if its clock is slow). But there is one value for the amount of clock error that makes all four spheres meet perfectly, and so a few algebraic computations can determine the necessary adjustment. Thus, even a simple receiver, with an electronic clock that is no more complicated or expensive than an ordinary digital wristwatch, can be synchronized with the atomic clocks whizzing past high in the sky.

**Pseudo-this, Pseudo-that**

Next, the military engineers who designed the system needed to decide how exactly to transmit the signals from the GPS satellites. They borrowed a technique that had been employed, strangely enough, by astronomers, among others, since the 1950s. Those scientists had been examining other planets by sending out radar pulses from their giant radio telescopes at what might have seemed to be random moments, but they were in reality following a carefully formulated code. The astronomers called these special cadences pseudo-random sequences. With them the researchers were able to measure the time delay in the weak radar reflections from the surface of a distant planet by finding the instant when the received signals and the transmitted pseudo-random sequence seemed to match most closely. In essence, the radar astronomers found the travel time (and hence the range to the radar target) by measuring when the two signals were most closely correlated.

Noting the astronomers' success, the military engineers opted to use similar pseudo-random sequences for their new space-based positioning system. They decided, however, that the GPS satellites would emit high-frequency radio waves continuously rather than beam discrete radar pulses down to the earth. The use of pseudo-random sequences to code the radio emissions offered many advantages. One that would be greatly appreciated by boaters and hikers years later was that it allowed inexpensive GPS receivers to be built. All satellites could then transmit on the same frequency without creaking a garbled mess of radio interference. Because each GPS satellite would transmit a unique code, an inexpensive, single-frequency radio receiver could easily separate the different signals.

The final decision that the military designers had to make concerned where to put the satellites. Nearly all space-bound hardware is placed in one of two types of orbit—either circling relatively close to the earth (in so-called low-earth orbit) or fixed at some 36,000 kilometers above the equator in a 24 hour-long geosynchronous orbit. Low orbits would cost relatively little for each launch and would demand only modest power from the satellites' transmitters because they would not have to broadcast their signals for any great distance. But such placement
would necessitate that hundreds of separate satellites be swarming around the planet to provide global coverage. Lofty geosynchronous orbits, on the other hand, would require far fewer satellites, but each would have to carry a more powerful transmitter, and these signals would have difficulty reaching the earth's polar regions.

The GPS planners chose a compromise solution, launching the satellites into orbits that were neither particularly low nor high; the satellites were set to orbit at an altitude of about 20,000 kilometers. At that altitude, 17 satellites would be sufficient to ensure that at least four of them—the minimum number needed to fix a position—would always be available from any location on the earth's surface. The final configuration adopted for the GPS has 21 primary satellites and three spares in orbit.

Selective Service

Because the U.S. defense forces intended to achieve a tactical advantage with the new satellite navigation system, from the outset they encoded the radio emissions to prevent adversaries from also gaining the ability to determine locations precisely. But the Department of Defense anticipated permitting ordinary people to use the GPS, at least in a coarse fashion. So the system's designers faced the question of how to limit civilian accuracy while still allowing the U.S. military to use the system to its full potential. There were several ways this dual operation could be accomplished. One method was to transmit incorrect information to unauthorized parties about exactly when the satellites had sent their signals. The GPS timing could be forced off slightly by altering the satellites' atomic clocks according to a specific code. Such "dithering" of the clocks appears to be what the Department of Defense has employed to keep the GPS secure, a procedure they term "selective availability." The modified signals allow all citizens to locate themselves reasonably well; navigational fixes will be off by no more than 100 meters. Military receivers that are equipped to interpret the classified code can readily work out a more refined position by removing the clock errors that have been added.

Civilian scientists and engineers interested in the GPS did not take long, however, to work out ways to get around the limitations of selective availability. Soon after the first group of these navigation satellites was launched, scientists had managed to find ways to reduce GPS errors—sometimes to as little as a few millimeters—achieving a level of accuracy that was many thousands of times finer than the system's military designers had thought possible. The first demonstration of such capability, by Charles C. Counselman III of the Massachusetts Institute of Technology and his colleagues, was performed rather unceremoniously in the parking lot of Haystack Observatory in Westford, Mass., during the fall of 1980.

To achieve a substantial improvement in accuracy, the scientists at Haystack and M.I.T. needed to correct the errors in the atomic clocks on the GPS satellites. The technique they employed was in fact quite simple: at a fixed point on the ground, they measured signals from several satellites. Knowing the exact location of the receiving antenna and the satellite positions, the scientists could then easily compare the pseudo-ranges (which they had measured) with the actual ranges (which they could calculate). The difference between the two numbers represented the error in the satellite clock, plus any inaccuracy in the clock used by the receiving equipment on the ground. The procedure of examining several satellites simultaneously allowed the scientists to determine the clock error on the ground, and hence they could work out exactly how much each of the spaceborne timekeepers was off.

The same method can be employed for circumventing selective availability today. The amount of clock dithering can be determined at a fixed ground station, and the corrections can be broadcast by radio. Mobile GPS apparatus operating nearby can use the information to calculate accurate locations. This scheme of "differential GPS" offers people outside the U.S. military the means to work out their whereabouts to within about a meter using surprisingly inexpensive equipment. (More specialized GPS receivers can achieve precision to about a centimeter.) There are currently a multitude of sources for differential GPS corrections. Many of them, curiously, are run by the U.S. government itself. The Federal Aviation Administration, for example, is starting to provide these services for aircraft. The U.S. Coast Guard, too, transmits corrections near major harbors. In addition, several commercial companies sell GPS corrections for most parts of the U.S. and for some other regions of the world as well.

The widespread availability of differential GPS has sparked considerable debate as to why the U.S. military continues to spend money to encode the GPS during peacetime, forcing other branches of government to expend yet more resources to decode the errors and broadcast the results. Ironically, during two recent military
operations, the Persian Gulf War and the occupation of Haiti, the Department of Defense turned off the security features of the GPS. They did so because there was not enough classified GPS equipment to go around, whereas civilian models were relatively easy to come by. (Many U.S. troops obtained this equipment, if in no other way, by telephoning home and purchasing GPS sets with their credit cards. The U.S. military had counted on their adversaries' lack of GPS-guided missiles and poor access to mail-order shopping.)

Moreover, the Russian government is now in the final stages of completing a satellite positioning system called GLONASS (for "Global Navigation Satellite System") that is largely similar to GPS. The Russian navigation system, however, does not encode its broadcasts, and thus anyone with the proper equipment can use it to full advantage. The existence of unencoded GLONASS, along with the widespread availability of GPS corrections, seems to negate any military advantage that might have accrued from purposeful degradation of the satellite clocks. A study recently conducted by the National Academy of Sciences has advised that selective availability is ineffective and should be discontinued. But so far the Department of Defense is still dithering--on the earth and in space.

**Where Next?**

With each passing week, people seem to find clever new applications for satellite positioning. Meteorologists are measuring the delays in GPS signals caused by the atmosphere to aid in weather forecasting. Farmers are using this equipment to survey the condition of each square yard of their fields so that they can distribute fertilizer most effectively. And, reasonably enough, the GPS is more and more helping to guide shins, airliners, helicopters, satellites and even passenger cars. Experimental systems carried in backpacks may eventually lead blind people about. Indeed, the commercial applications now far outnumber the military uses of the system, and by the turn of the millennium the sale of GPS services should bring about $1 billion into the U.S. economy every year. With unprecedented speed, what was born as a military system has become a national economic resource. In this rapidly changing world, one must seriously wonder: Who should control the GPS?

**Catching the Waves**

Geophysicists have applied the GPS since the mid-1980s to help monitor the slow but relentless deformation of the earth's crust in geologically active regions--changes that can eventually cause the ground to rupture in an earthquake. For such investigations, they seek a maximum in precision from the GPS and often employ a technique called carrier tracking. Compared with differential GPS (which can locate a position to within about a meter or so), carrier tracking allows locations to be determined to within a few millimeters.

Carrier tracking gets its name from the satellite broadcasts that convey GPS signals on a set of so called radio carrier waves. It works by determining which part of the radio wave strikes the antenna at a given instant--the "phase" of the received emission. Like an ocean swimmer sensing whether he or she is positioned at the crest of a wave, in a trough or somewhere in between, carrier tracking evaluates where on the 19-centimeter-long GPS radio waves the receiving antenna sits.

Carrier tracking allows a resolution that is a tiny fraction of a wavelength. The primary difficulty is in determining which of many identical waves the antenna is sensing. There are, however, a number of ways to overcome such ambiguity. The simplest is to track the carrier phase from several satellites simultaneously. If, for example, one found that the receiving antenna was located at the start of the waves (at zero phase) sent from three different satellites, there would be a limited number of spots where that coincidence was possible (solid points on diagram). With enough satellites, allowable points are spaced over a meter apart. Hence, by knowing the approximate position of the antenna (using the differential GPS technique), one can determine which of the points located by carrier tracking marks the correct location.

In practice, carrier tracking proves to be a rather delicate undertaking. The passage of the waves and the motion of the satellites need to be accounted for. Some uses of carrier tracking, such as the aircraft landing system developed by Bradford W. Parkinson and his colleagues at Stanford University, require special hardware to ensure the integrity of the navigation solutions. These efforts are made more difficult by a security feature introduced by the Department of Defense called anti-spoofing. Like the encoding of clock errors (called selective availability), anti-spoofing makes many marvelous high-precision civilian applications of the GPS harder and more costly to accomplish.
DIAGRAM: CLOCK ERROR in the receiving electronics typically causes the distance measurements to the GPS satellites to be somewhat incorrect. With such inaccuracies, the corresponding spheres of position (thick lines) will not intersect neatly at a single point. Adjusting the receiver's clock slightly forward or back does, however, correct the ranges and allows all spheres to meet precisely (thin lines). This method appears here to require measurements from only three satellites, but in three dimensions, four satellite ranges are necessary.

DIAGRAM: ATMOSPHERIC LAYERS alter the GPS signals and can introduce significant errors. The most important effects arise as the radio waves pass through the earth's charged ionosphere and water-laden troposphere. Whereas the radio wave fronts (inset) tend to stretch out in the ionosphere, they bunch together in the troposphere. Because these disturbances to the GPS signals can be measured with a fixed receiver, scientists at the National Weather Service can now measure atmospheric water content this way.

GRAPH: PSEUDO-RANDOM SEQUENCES are broadcast from the GPS satellites at known times. The delay in the arrival of the radio emissions is found by comparing the noisy signal received (red) with versions of the known sequence (blue) that are shifted in time. A high correlation between the code sequence and the signal (right) indicates the time lag between transmission and reception of the signals.

PHOTO (COLOR): SPHERES OF POSITION show the geometric basis for the operation of the Global Positioning System (GPS). By receiving coded broadcasts, a person on the earth can determine his or her distance to each of several satellites that orbit in precisely known patterns. Each distance measurement coincides with a set of possible locations that form an imaginary sphere (purple or green) centered on the satellite emitting the signal. The intersection of several spheres with the surface of the earth marks the person's exact location.

PHOTO (COLOR): DIFFERENTIAL GPS circumvents the clock errors imposed for military security. A fixed receiver at a known location determines the clock errors in the satellite signals (red) and broad casts the appropriate corrections (blue) to mobile receivers in use nearby. This method can decrease the uncertainty in GPS positioning from 100 meters to as little as one meter.

PHOTO (COLOR): BLIND NAVIGATION is no longer restricted to ships in fog. Visually impaired people may be able to use GPS to get around outdoors. This experimental application, and others, would be far easier without the confounding effects of military security measures.

PHOTO (COLOR): DEFORMATION OF THE EARTH'S CRUST in geologically active areas such as the Tien Shan of central Asia can be measured using precise GPS surveys of benchmarks. The GPS technique thus serves as a research tool to help monitor the accumulation of strain that can eventually cause devastating earthquakes.

ILLUSTRATION

Further Reading


GPS WORLD. Bimonthly magazine published by Advanstar Communications, 859 Williamette Street, Eugene OR 97401.


The University Navstar consortium site on the World Wide Web at http://www.unavco.ucar.edu/

by Thomas A. Herring
THOMAS A. HERRING is an associate professor in the department of earth, atmospheric and planetary sciences at the Massachusetts Institute of Technology. During his undergraduate studies at the University of Queensland, he spent his summers working underground in Australian mines and wrestling his way through the jungles of Papua New Guinea. These experiences convinced Herring that it was safer to pursue graduate studies in front of a computer at M.I.T., where he learned how to apply precise measurement systems to geophysical problems. In 1983 he received his Ph.D. and then joined the Harvard College Observatory. Herring remained there for six years before returning to M.I.T. After many years of respite from the rigors of fieldwork, Herring once again finds himself involved in field studies in such far-flung places as the central Asian republic of Kyrgyzstan and the Parana River in Argentina.