how it works

ALMOST THREE MILLION PEOPLE ARE LOOKING FOR SIGNS OF EXTRA-TERRESTRIAL INTELLIGENT LIFE.

Is anyone talking?

By Gabe Moretti, Technical Editor

April 2001 marks the second anniversary of one of the largest, if not the largest, distributed-computing experiments, ever attempted. SETI@home is one of a number of projects searching for evidence of extraterrestrial intelligence, but it is unique in its approach. You need a lot of computing power to filter and analyze data when you're looking for something whose characteristics you can only guess about in the enormous sampling space that is the entire universe. One approach is to obtain a number of supercomputers and a large supporting staff to perform the data-analysis task in a centralized data center. This method requires a significant initial investment and large operational costs due to the specialized nature of the computing system. For scientists searching for little green men, the probability of obtaining any significant funding from either the government or a foundation is so close to zero, it is impractical to even consider it.

David Anderson, the project director of SETI@home, and Dan Werthimer, its chief scientist, have implemented a new approach. (See **sidebar** "Two Davids, Captain Picard, and the Planetary Society get together" for the project history.) Their method exemplifies the creative employment of underused resources, and, if successful, it will be a triumph of ingenuity over the cynicism about funding basic science. It uses an otherwise redundant antenna at Arecibo, Puerto Rico, and the idle cycles of PCs and workstations throughout the world to perform most of its required tasks. The system comprises four sequential phases: data capture, formatting, analysis, and validation.

When conducting a SETI (search for extraterrestrial intelligence), in addition to computing power, you need to establish the characteristics of what you are looking for, and you need to obtain an instrument powerful enough to guarantee that you capture the physical phenomena of the object of your search. Because helium is the most abundant element in the universe, it is logical to assume that in choosing a likely common "language" in which to communicate, a civilization would transmit signals around the 1.42-GHz frequency, corresponding to the spectrum line of helium. SETI@home looks at a 2.5-MHz-wide band cen-

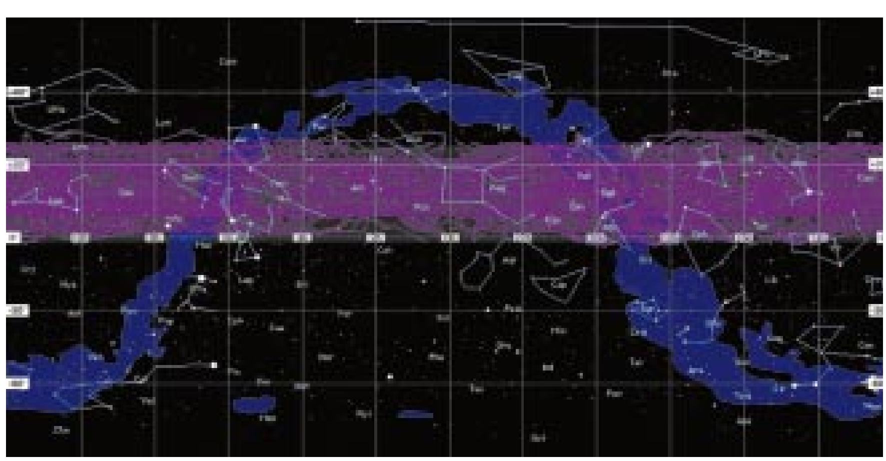
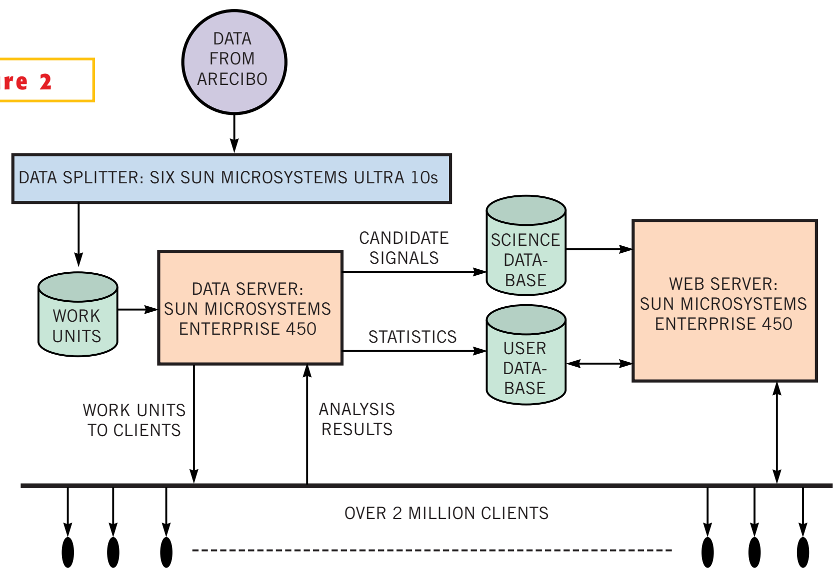


Figure 1 The path of the Arecibo antenna has covered a significant portion of the Milky Way visible from the Northern Hemisphere.

tered on the helium line. The receiving instrument is the Arecibo radiotelescope. The radiotelescope is a spherical structure constructed in a natural bowl-shaped valley in Puerto Rico. The receiving instrument is suspended 700 feet above the center and contains two antennas—a primary antenna and a secondary antenna—that make the structure symmetric and easier to focus and that provide redundancy in case of failure. The use of the primary antenna costs more than $2000 an hour, but the secondary antenna just goes along for the ride. It tracks the movements of the primary antenna with a 6° angle of separation. SETI@home is looking for signals coming from our home galaxy, the Milky Way, but it does not have to search in any particular area at any particular time. **Figure 1** shows the area of the sky that it has searched so far.

Figure 2



The modular architecture of the server provides opportunities to increase the number of required functions without impacting the total system.

The project is looking for three types of signals: a narrowband pulse; repeated pulses; and chirp signals, whose frequency diminishes over time due to the Doppler effect. You must remember that in the universe nothing is stationary. Every object is in motion, and two objects move with respect to each other. Due to the Earth's rotation, the Arecibo antenna sweeps an object in 15 seconds from the time it encounters the target limb to the time it leaves the opposite limb. So, a true signal would have a predictable Gaussian pattern.

The Earth is immersed in radio frequencies, most of which people generate. Therefore, the captured signal stream must be frequency-shifted, filtered, and sampled. These functions are performed at Arecibo, and the "clean" signal is then sampled using a 2-bit algorithm, resulting in 2.5M samples/sec. The data is recorded on 35-Gbyte digital linear tape. Each tape can store 16 hours of data. When the tape is full, a light in the control room at Arecibo alerts a technician to mount a new tape. When a box-worth of tapes has been recorded, the tapes are shipped by Federal Express to the SETI@home offices in the Space Sciences Laboratory on the University of California—Berkeley campus.

At the university, a 2-D formatting process, employing a half-dozen Sun Microsystems Ultra 10 machines, divides the 2.5-MHz stream into 256 slices of 10-kHz each. Each slice is then divided into segments representing 107 seconds of recording time. Two consecutive segments will have a 20-second overlap to ensure that, if a signal is present, its entire Gaussian profile will be contained in one segment. (Remember that each Gaussian profile is 15 seconds.) Each segment contains slightly more than 350 kbytes of data, a length that you can transmit over telephone lines at 28.8 kbps in about two minutes. Each segment is a "work unit," and the work units are the elements that the system distributes throughout the world for analysis. **Figure 2** shows the architecture of the server portion of the system.

The architecture of the analysis portion of the system is an open-ended server-client distributed system. The client is a "screensaver" program that uses the spare cycles of a computer to perform data analysis and reduction. Almost 2 million computers are currently running the screensaver. The screensaver only requires hardware with the ability to perform floating-point arithmetic and a connection to the Internet to report results and obtain new work units.

The server architecture contains work-unit storage of approximately 500 Gbytes, which stores the result of the data-formatting phase; a data server that communicates with the clients, sends out work units, and receives results; a science-database server; a user-database server; and a Web server. The server machines are Sun Microsystems Enterprise 450s, and the data is stored using the Informix relational-database engine. The database engine and all of the Sun Microsystems equipment have been donated to the project. The data server stores each client result in the science database for further processing. These results are the "candidate signals" that scientists will further analyze. The system maintains operational statistics in the user database and groups them according to individual user, team, country, operating-system type, platform, and location.

You can appreciate the fact that SETI@home must be sure that, if it ever finds a signal, it is the genuine article. One of the ways to improve the quality of results is to perform redundant analysis. This practice is not only good science, but also the result of pragmatic observations made during the course of the project. Most client machines are Intel-based PCs, PCs based on Intel clones, or Apple computers. Despite efforts by the IEEE to standardize floating-point arithmetic operations, these machines do not produce standard floating-point results. So the server sends each work unit to three clients for analysis. The final phase of the system takes the candidate results and performs RFI rejection on them to screen out spurious human-generated signals. The surviving candidates are then grouped by their work-unit ID, and the server performs a redundant result comparison using the time-tested voting method: If two results match, the candidate is saved; otherwise, it is rejected. Finally, the system checks for repeat signals coming from the same coordinates in the Milky Way. In the two years of operation, no one has found any, but two years is a short time.

As you might expect, the science database is getting quite large. The next modification to the architecture will be to clone the database by modifying the validation phase. The science database will be analyzed, redundant candidates will be removed, and a master candidate database established. The project scientists will then segment and clone this master database so that a complete historical record can facilitate the repeat-signal-checking process. Anderson hopes that his organization will soon obtain funds to expand the search by using a radiotelescope in Australia, so that it can study the other half of the Milky Way—the one visible from "down under." A dozen scientists and 2 million volunteers from 108 countries know we are not alone, we just have not yet found our neighbors or have not yet learned how to say "hello."□

1-303-652-0480,
1-303-652-0479,

TWO DAVIDS, CAPTAIN PICARD, AND THE PLANETARY SOCIETY GET TOGETHER

1986,
200,000.
2.8
$50,000
1994,
Star Trek: Insurrection,
25
$50,000
11
1999,
$300,000,
1995,
&
1999
5000
1998,
1999,