

The following paper was published (1997) in:

"Astronomical and Biochemical Origins and the Search for Life in the Universe", Proc. of the Fifth Intl. Conf. on Bioastronomy = IAU Colloq. No. 161, eds. C.B. Cosmovici, S. Bowyer, and D. Werthimer (Publisher: Editrice Compositori, Bologna, Italy)

A new major SETI project based on Project Serendip data and 100,000 personal computers

W. T. Sullivan, III (U. Washington), D. Werthimer, S. Bowyer, J. Cobb (U. California, Berkeley), D. Gedye, D. Anderson (Big Science, Inc.)

ABSTRACT

We are now developing an innovative SETI project, tentatively named *seti@home*, involving massively parallel computation on desktop computers scattered around the world. The public will be uniquely involved in a real scientific project. Individuals will download a screensaver program that will not only provide the usual attractive graphics when their computer is idle, but will also perform sophisticated analysis of SETI data using the host computer. The data are tapped off Project Serendip IV's receiver and SETI survey operating on the 305-meter diameter Arecibo radio telescope. We make a continuous tape-recording of a 2 MHz bandwidth signal centered on the 21 cm HI line. The data on these tapes are then preliminarily screened and parceled out by a server that supplies small chunks of data (50 seconds of 20 kHz bandwidth, a total of 0.25 MB) over the Internet to clients possessing the screen-saver software. After the client computer has automatically analyzed a complete chunk of data (in a much more detailed manner than Serendip normally does) a report on the best candidate signals is sent back to the server, whereupon a new chunk of data is sent out. If 50,000-100,000 customers can be achieved, the computing power will be equivalent to a substantial fraction of a typical supercomputer, and *seti@home* will cover a comparable volume of parameter space to that of Serendip IV.

Introduction

Science, although almost totally supported by public funds, has traditionally been carried out in laboratories and observatories not open to the general public. In an era when the public's support of science is wavering, this modus operandi may be self-defeating and requires re-examination. The goal of the present SETI project, tentatively named *seti@home*, is (a) to do good science, and (b) to do it in a way that engages and excites the general public. This is a chance to educate participants about how science works, as well as to give them reliable information about SETI (as opposed to, for example, the film *Independence Day*). In the end the scientific community can only profit if the public better understands the scientific enterprise.

Once operational, *seti@home* will:

- be the first time ordinary computer users will be able to participate in a major scientific experiment.
- be the largest distributed computation ever undertaken.
- provide for each participant the slight but captivating possibility that his or her computer may be

instrumental in the detection of another technical civilization in the Milky Way.

Project Architecture and Data Flow

Seti@home is a "piggyback" survey based on the Serendip IV survey, which itself is a piggyback survey operating on the 305-m Arecibo telescope. Serendip IV, which will begin operations on the newly upgraded Arecibo dish in 1997, is described (along with its predecessor) by Bowyer et al. (1997) elsewhere in this volume. The basic idea is that a separate Serendip receiver and data processor "rides around" on the Arecibo feed platform as normal radio astronomy is carried out. The sky visible to Arecibo is thus surveyed in a pseudo-random fashion, and in fact any given patch of sky is typically revisited every 3-6 months. These revisits are critical for discriminating against manmade radio-frequency interference (RFI).

The overall architecture of seti@home is presented in Figure 1. At Arecibo seti@home will tape-record at baseband a small portion of Serendip's total bandwidth of 183 MHz; 2 MHz will cover the 3 possible velocity rest frames of heliocentric, galactocentric, and cosmic background radiation. This band will have been down-converted from an observing frequency centered on the 21 cm hydrogen line at 1420 MHz, ideal in terms of SETI strategy and freedom from manmade interference (RFI). With one-bit sampling at the Nyquist rate, this is a data-recording rate of 0.5 MB/sec, or one Exabyte Mammoth DAT (25 GB capacity) every 11 hours, which is 500 per year (for an expected 70% observing efficiency). These tapes are Fed-Ex'ed to the Big Science server where they are validated, archived, and supplied with indexing parameters, in particular sky positions and times. Only a portion of the data is ever analyzed, in the first instance whatever is necessary to keep customers supplied (see below). This portion of the data will be chosen so as to be at the lowest possible galactic latitudes, i.e., in the direction of the galactic plane as in the "Milky Way strategy" put forward by Sullivan and Mighell (1984).

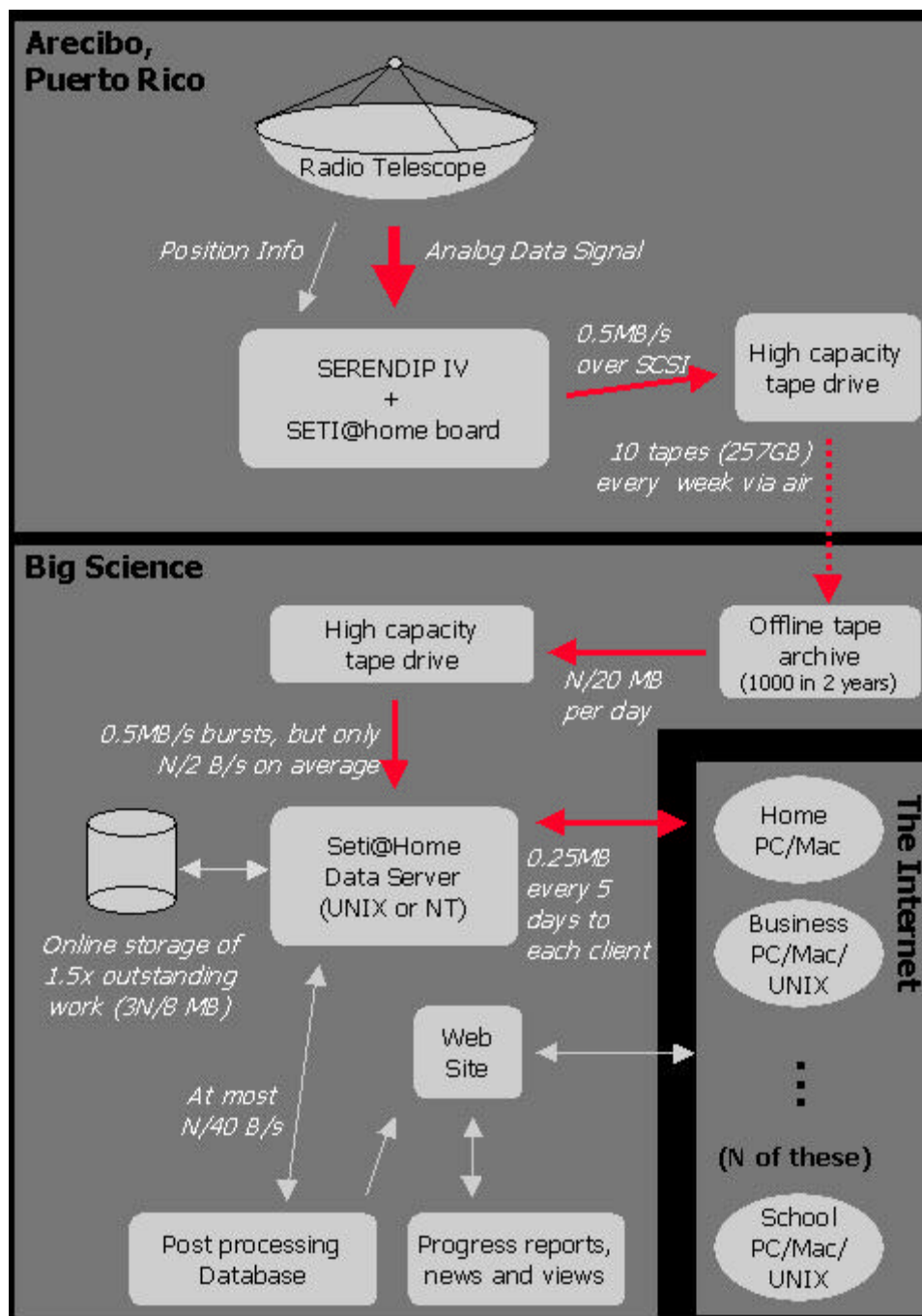


Figure 1. A schematic diagram of the architecture and data flow of seti@home.

Note that "N" refers to the number of computers currently participating in the analysis.

Even if no signal appears on the first look at a given sky position, sky positions that are (by chance) repeated are given the highest priority for analysis because they allow one to combat RFI and to treat the possibility of interstellar scintillations. Candidate signals are assigned higher IQ's (Interest Quotients) according to how many times they repeat and how distinct they look from possible RFI in terms of their fitting the beam pattern and exhibiting a non-local Doppler drift.

Each customer will work at any time on 0.25 MB of data, which represents 50 seconds of a 20 kHz signal. These 20 kHz bandwidths will be created from the original 2 MHz data at the server by lookup-table FFT's. This 50 sec of drift data corresponds to a sky swath of size 6' x 25' (4 independent

beams, essential for RFI discrimination). This means that each person has a 50 sec x 20,000 Hz array as his/her basic data set; code and data storage requirements to analyze this chunk are estimated at 3-4 MB. On the receiving end a 0.25 MB chunk will require 1.3 sec on an incoming T1 line of 190 kB/s, or 2.3 minutes on a 14.4K baud line (sufficiently short not to be discouraging to customers on phone lines). Upon completion (typically after several days) of the data analysis for each chunk, a short message reporting candidate signals is presented to the customer and also returned to Big Science and to the University of Washington for post-processing (see below).

On the server end, a T1 line can handle 90 GB per week and thus 360,000 customers could be serviced weekly (or 180,000 twice per week) with one outgoing line. This 90 GB represents 30% of a fully-recorded week at Arecibo, or 43% of a more realistic week (at 70% efficiency).

Signal Analysis @home

The uniqueness and advantage of seti@home from a SETI point of view is the depth of signal analysis that will be possible on each chunk of bandwidth x time (and sky position). (N.B.: The customer will not have any control over the nature of the search; the home computer's CPU cycles are simply being rented by an automatic program.) It is proposed to look for signals of widths ranging from 0.1 Hz (a 2×105 point FFT) to 2000 Hz. (Serendip now searches for lines of width 0.6 to 640 Hz; the partial redundancy between the two searches will be a useful cross-check.) All signals will be sought for each of 40 assumed distinct Doppler drift rates (during the 12 sec pass time through one beam); anything that has zero drift with respect to Puerto Rico will automatically be rejected as RFI, but all nonzero drift rates will be retained, especially those matching the earth's rotational and orbital acceleration (of order 0.5-1.5 Hz over 12 sec; the actual value at the time of observation will be calculated by the server and tagged on the data chunk). These data also allow searching for intermittent or pulselike signals with durations anywhere from 0.5 msec to 10 sec - this is one of the main advantages of seti@home over Serendip, which can only detect pulses with a spacing greater than 1.7 sec. Any signals lasting for less than a beamwidth might be burst-like RFI, but will be retained (albeit with lower weight). The most viable candidates, however, must also match the beam-shape, for this is the best criterion for distinguishing a real point-source signal from RFI. Thus a Gaussian beam correlation filter (with a width of 12 sec) will also be applied to all candidate signals. In summary, candidate signals will be reported back to the server on the basis of their exhibited signal/noise ratio for the best-fit bandwidth, timewidth, frequency channel, frequency drift, and beam shape matching.

Post-Processing

Post-processing at the University of Washington will build up a database of all candidate signals reported back from the customer computers. This database will be used to monitor progress with the survey, monitor data quality and the RFI environment at Arecibo, and decide if any changes in survey or operational procedures may be desirable. With the extremely high number of cases that we will be examining, we expect that RFI will often by chance behave in such a manner that it mimics an extraterrestrial signal. Thus it is extremely important that any high-signal/noise candidates observed once are only considered highly tentative until positively confirmed by other (random) passes (usually 3-6 months later) of the beam over the same patch of sky. The post-processing will monitor for these subsequent passes and compare the consistency of the characteristics of the candidate signals for each pass. In this way a set of best candidate signals will be accumulated, to be checked later with follow-up, dedicated Arecibo time.

Post-processing will also search for characteristic regularities of the reported signals as members of a

group. One example are frequency multiplets, namely pairs or n-tuplets of arbitrarily, but equally spaced narrowband signals. For the advantages of such signals for SETI, see Cohen (1994) and Cordes and Sullivan (1994).

It should not be forgotten that there exists the tantalizing possibility of serendipitously discovering, as a byproduct of this SETI analysis, a wholly new type of naturally occurring astrophysical phenomenon. The overall search analysis is the same except that natural signals could be broader than a single beam and are therefore that much trickier to distinguish from RFI. For candidates smaller than a beam, however, the distinction between an intelligent or natural origin can only come later after follow-up observations and analysis.

The Screensaver Itself

The program that runs on each client computer looks and behaves like a captivating screensaver. It runs only when the machine is idle, and the user can choose from several different colorful and dynamic "visualizations" of the SETI process. Some of these visualizations will look technical, some will look abstract, and some will look decidedly artistic. We will provide a core set of visualizations, as well as a plug-in mechanism so that others can easily be added. Standard screensaver modes will include (1) a map of the world showing the location of all machines currently participating in the project, (2) a map of the sky showing what areas have been covered by the survey and the location of the patch of sky currently being analyzed (with the option of viewing classical mythological figures for the constellations - one might, for instance, see that one's patch is in the armpit of Orion!), (3) colorful, changing patterns that correspond to the Fourier transforms currently being undertaken, and (4) "straight" graphs showing results of the currently evolving data analysis.

Closing

As of September 1996 seti@home is only a proposal, but we feel certain that it will happen soon. The potential customers for seti@home include astronomy and space enthusiasts (e.g., members of The Planetary Society), science fiction fans (e.g., "Trekkies"), Internet adventurers ("netheads"), science teachers and their students, and other science & technology enthusiasts. All we need to make this global participatory project a reality is sponsorship by a visionary, high-tech corporation. The project is good science and it will generate great publicity and goodwill for any sponsor, as well as for science in general and bioastronomy in particular. Any takers?

REFERENCES

- Bowyer, S., Werthimer, D., Cobb, J., and Lampton, M. 1997, "Twenty years of SERENDIP, the Berkeley SETI effort: Past results and future plans", this volume.
- Cohen, N., and Charlton, D. 1995, "Polychromatic SETI", pp. 313-22 in *Progress in the Search for Extraterrestrial Life*, ed. G.S. Shostak (San Francisco: Astronomical Soc. of the Pacific).
- Cordes, J., and Sullivan, W.T., III 1995, "Astrophysical coding", pp. 325-42 in *Progress in the Search for Extraterrestrial Life*, ed. G.S. Shostak (San Francisco: Astronomical Soc. of the Pacific).
- Sullivan, W.T., III, and Mighell, K.M. 1984, "A Milky Way search strategy for extraterrestrial intelligence", *Icarus* 60, 675-84.