Bioastronomy 2002: Life Among the Stars
IAU Symposium, Vol. 213, 2004
R.P.Norris and F.H.Stootman (eds.)

Three Years of SETI@home: A Status Report

Eric J. Korpela, Jeff Cobb, Steve Fulton, Matt Lebofsky, Eric Heien, Eric Person, Paul Demorest, Robert Bankay, David Anderson & Dan Werthimer

Space Sciences Laboratory, University of California, Berkeley, CA 94720-7450, U.S.A.

Abstract. The SETI@home project has recently completed its third year of active data analysis. Over 4 million volunteers have joined the search, providing a combined total of over 1 million CPU-years of processing power. SETI@home performs a sensitive search for extraterrestrial signals in a 2.5 MHz band centered on 1420 MHz. SETI@home searches a wide parameter space including 14 octaves of signal bandwidth and 15 octaves of pulse period with Doppler drift corrections from -50 Hz/s to +50 Hz/s. We will briefly describe the SETI@home project and the algorithms used in the SETI@home client. We will describe the post-processing methods we use to reject RFI and select candidate signals from the nearly 4 billion "hits" returned by SETI@home clients.

1. Observing Methodology

The UCB SETI searches use the 1420 MHz line feed on Carriage House 1 at the National Astronomy and Ionospheric Center's 305 meter radio telescope in Arecibo, Puerto Rico. This unique arrangement allows observations to be conducted without interference with other uses of the telescope. This results in two main modes of observation. If the primary observers feed is stationary or stowed the beam scans across the sky at the sidereal rate. If the primary observer's feed is tracking a position on the sky, the beam scans the sky at twice the sidereal rate. At twice the sidereal rate, the beam width corresponds to a 12 second beam transit time (Korpela et al. 2001). Figure 1 shows the path of the telescope beam over the course of 15 hours. Since the start of the project, the telescope has covered about 90% of the sky visible from Arecibo.

The time domain data for the sky survey is recorded as follows: first, a 30 MHz band from the receiver is converted to baseband using a pair of mixers and low pass filters. The resulting complex signal is digitized and then filtered to 2.5 MHz using a pair of 192 tap FIR filters in the SERENDIP IV instrument. (Werthimer et al. 1997) One bit samples are recorded on 35 GByte DLT tapes (one bit real and one bit imaginary per complex sample). These tapes are shipped to Berkeley for use in the SETI@home program.

In Berkeley, the data from the tapes are split into work units of duration 107 seconds and bandwidth 9766 Hz. These work units are shipped over the Internet

https://doi.org/10.1017/S0074180900193635 Published online by Cambridge University Press

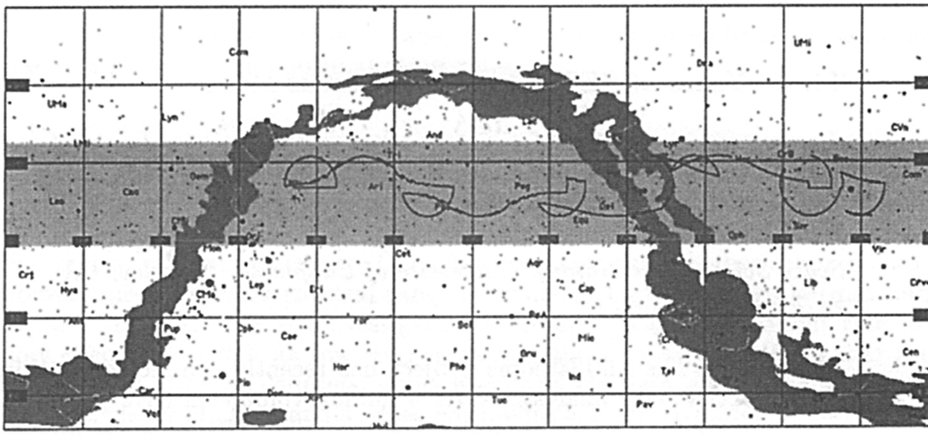

Figure 1. The path of the telescope beam over 15^h on August 5, 2001.

to the computers of volunteers running the SETI@home client program. Details of the server and distribution method can be found in Anderson et al. (2002).

2. The SETI@home Client Program

After receiving a work unit, the client performs a baseline smoothing on the data to remove any wide-band $\Delta\nu > 2$ kHz features. This prevents the client from confusing fluctuations of broadband noise (due in part to variations in the hydrogen line emission as the field of view transits the sky) with intelligent signals. The client then begins the main data analysis loop.

At the start of each passage through the loop, the data is transformed into an accelerated frame of a given Doppler drift rate. The drift rates at which the client searches the data for signals vary from -10 Hz/s to +10 Hz/s (accelerations expected on a rapidly rotating planet) in steps of 0.0018 Hz/s. The client also examines the data at Doppler drift rates out to ± 50 Hz/s (accelerations of the magnitude that would arise from a satellite in low orbit about an earth-like planet), but at a more coarse step of 0.029 Hz/s. A signal from an alien world would most likely have a negative drift rate (as the accelerations involved would be away from the observer).

Despite this, we examine both positive and negative drift rates for the purpose of statistical comparison and to leave open the possibility of detecting a deliberately chirped extraterrestrial signal.

At each drift rate, the client searches for signals at one or more bandwidths between 0.075 and 1 221 Hz. This is accomplished by using FFTs of length 2n (n=3, 4, ..., 17) to transform the data into a number of time-ordered power spectra. To avoid repeating work, not all bandwidths are examined at every Doppler drift rate. Only when the change in drift rate becomes significant compared to $(1/\Delta\nu^2)$ does the program compute another FFT of a given length. Therefore, 32K-point transforms are performed one quarter as often at 64K-point transforms.

https://doi.org/10.1017/S0074180900193635 Published online by Cambridge University Press

The transformed data is examined for four types of potential signals:

1. Spikes: any FFT bin that exceeds 22 times the mean noise power.
2. Gaussians: a signal exceeding 3.5 times the mean noise power which matches (χ^2 <threshold) the time profile expected as a source transits the field-of-view (beam) of the telescope.
3. Triplets: grouped events evenly spaced in time in which all events exceed $7.75\times$ the mean noise power and occur within the beam transit timescale.
4. Pulses: signals detected using a modified fast folding algorithm. The folding algorithm divides the data into chunks of duration equal to the period being searched and co-adds them to improve signal-to noise ratio. A computed threshold selects signals below a set probability level.

More details of the client algorithms can be found in Korpela et al. (2001).

3. Candidate Identification

Any potential signals detected are returned to Berkeley and stored in a database. Each work unit is processed by multiple client machines. The results are compared to reduce the probability of error or deliberate introduction of false signals. Signals at frequencies or pulse periods corresponding to known continuous RFI are removed from further consideration.

We then identify potential candidates for further examination. We define a candidate as a group of signals that satisfy all of the following criteria:

1. The signals are from the same position on the sky, within some positional tolerance, typically of order the beamwidth.
2. The signals were detected at significantly different times. This time separation should be large compared to the typical duration of transient RFI.
3. The signals are within the same barycentric frequency window. The width of this window can be varied. A small window (± 50 Hz) is used to search for signals that have been corrected for Doppler drifts at the transmitter. A larger window can be used to detect those that have not been corrected for Doppler drift. For pulsed signals, a pulse period window may also be used to detect signals with a constant pulse period.
4. The signals are above a signal score or power threshold. This threshold is chosen to keep the number of candidates at a level that shows the underlying distribution of random noise detections (typically a few hundred thousand candidates).

We assign each candidate a score proportional to its probability of arising due to random noise. As an example, Fig. 2 shows the score distribution of candidates consisting of multiple "pulse" signals, lower scores indicate low probability of arising due to random noise. The vertical features at harmonics of 25 Hz are pulsed RFI, which can easily be removed. The remaining distribution is fairly noise-like.

In order to allow candidates of different types to be compared, the scoring algorithm should be independent of signal type, frequency window, and power threshold. In the near future, we intend to examine coincidence between candidates and celestial objects. We may modify the scoring algorithm to include

https://doi.org/10.1017/S0074180900193635 Published online by Cambridge University Press

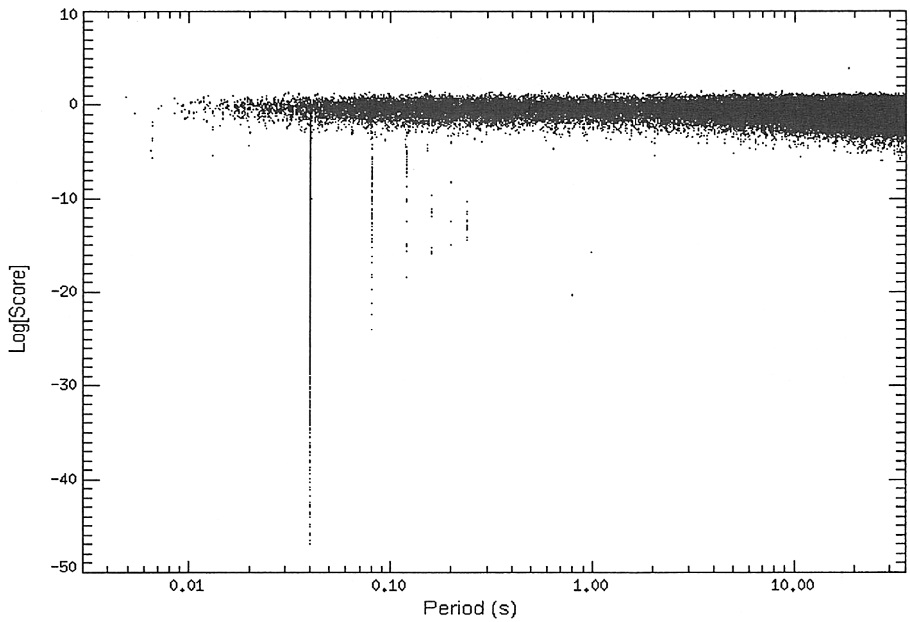

Figure 2. Score distribution vs. period for pulse candidates.

the probability of coincidence. We maintain a public archive on our web site (http://setiathome.ssl.berkeley.edu/) showing some of the best candidates.

An examination of the existing candidate lists indicates that many may still be due to RFI. We intend to run our signals through the SERENDIP RFI rejection suite (see Cobb et al. 1997) in order to remove faint RFI. We have proposed for dedicated telescope time at Arecibo to reinvestigate our top candidates.

Acknowledgments. SETI@home is primarily funded through private donations. We especially would like to thank The Planetary Society, The SETI Institute, Sun Microsystems, Network Appliance, The University of California, and The Friends of SETI@home for their generous support. We would also like to thank the staff NAIC Arecibo observatory. NAIC is operated by Cornell University under a cooperative agreement with the National Science Foundation.

References

Anderson, D. P., et al. 2002, Communications of the ACM, in press

Cobb, J. et al. 1997, in Astronomical and Biochemical Origins and the Search for Life in the Universe, ed. C. B. Cosmovici, S. Bowyer & D. Werthimer, (Bologna: Editrice Compositori), 677

Korpela, E. J., et al. 2001, Computing in Science and Engineering, 3, 79

Werthimer, D. et al. 1997, in Astronomical and Biochemical Origins and the Search for Life in the Universe, ed. C. B. Cosmovici, S. Bowyer & D. Werthimer, (Bologna: Editrice Compositori), 711

https://doi.org/10.1017/S0074180900193635 Published online by Cambridge University Press