A New Era in Bioastronomy
ASP Conference Series, Vol. 213, 2000
G. Lemarchand, K. Meech (eds.)

2000ASPC..213..485C

SERENDIP IV: Data Acquisition, Reduction, and Analysis

Jeff Cobb, Matt Lebofsky, Dan Werthimer, Stuart Bowyer, Michael Lampton

Space Sciences Laboratory, University of California, Berkeley, California, CA 94720-7450, USA

Abstract. The SERENDIP IV SETI program is an ongoing radio sky survey conducted by the University of California at Berkeley. The observations are conducted remotely at the National Astronomy and Ionospheric Center in Arecibo, Puerto Rico. Here we discuss the data acquisition and analysis pipeline from the detection of narrow band signals to the construction of an ETI candidate database. Particular emphasis is placed on the data processing algorithms employed to reject radio interference and to search for the patterns that lead us to classify sets of signals as ETI candidates.

1. Introduction

SERENDIP IV was installed at the Arecibo Observatory in Puerto Rico in June of 1997 and with the aid of a dedicated receiver is carrying out a multiple year sky survey in the 1.4 GHz band. As in previous generations of SERENDIP, this survey is a commensal search. By piggybacking on other regularly scheduled observations, SERENDIP IV is able to acquire data around the clock.

The SERENDIP IV observing cycle consists of real-time data analysis by the SERENDIP IV instrument, near-time data transfer across the Internet, offline data reduction and candidate generation, and finally re-observation of the most interesting candidates. This is illustrated by Figure 1.

2. Real-Time Data Analysis

The SERENDIP instrument consists of forty spectrum analysis / postprocessing boards working in parallel. See Werthimer *et al.* (2000) for details on hardware design. Each of the forty boards utilizes dedicated hardware to perform a 4 million point Fast Fourier Transform on a different section of the band. The calculated power spectra enter the post-processing portion of each board where baseline normalization, coarse resolution spectra computation and event thresholding occur. Baseline normalization is achieved with a sliding eight thousand channel local mean boxcar. Thresholding is performed based on mean spectral power and signals that exceed the threshold are logged along with time, pointing coordinates, frequency, and power.

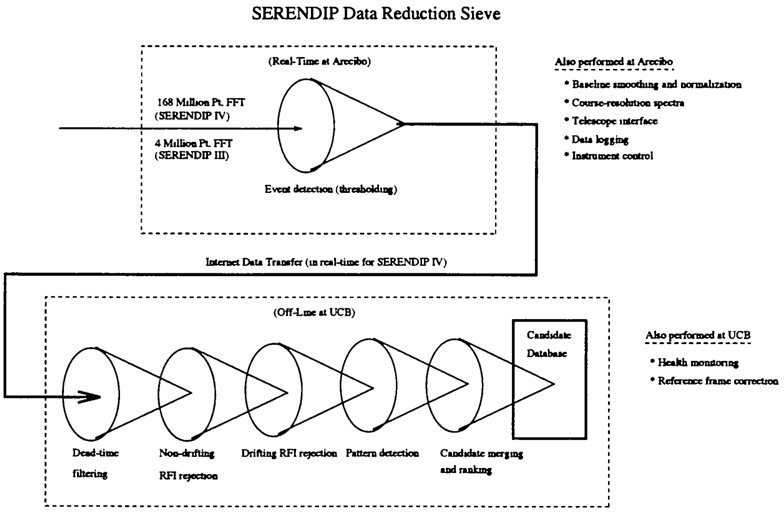


Figure 1. Each cone represents a data reduction step

3. Networked Data Acquisition

Signals exceeding the threshold are reported by the postprocessors to the control PC which logs these signals to a UNIX workstation via a dedicated ethernet. The workstation functions as a short term archive and data transfer machine. As data arrive from the PC, they are simultaneously sent across the Internet to UC Berkeley.

4. Off-Line Data Reduction

Off-line data reduction consists of telescope deadtime filters and a set of cluster analysis algorithms that filter radio frequency interference (RFI). Instrument health is also monitored at this stage by the detection of an artificial signal that is periodically injected into the instrumentation.

Telescope deadtime filtering removes events recorded during times when the slew rate of the telescope was incompatible with data acquisition and cluster analysis. SERENDIP receives the telescope's pointing coordinates every 5 seconds and interpolates the positions for spectra taken between each report. A too rapid slew rate leads to errors in this interpolation. Conversely, zero or very slow slewing makes it difficult to remove RFI. Data collected during both of these extremes are removed while those collected during normal slew rates are passed on to the RFI filters.

Passband, spread spectrum, and drifting RFI are removed during the RFI rejection stage. If a signal persists over more solid angle than a point source would allow, it is labeled as passband RFI and removed. If a spectrum exhibits many and widely spaced signals, it is labeled as being contaminated by spread spectrum RFI and is removed.

Even though typically 98% of all events are identified and rejected as RFI, only 1% of the band is lost during the filtering process. The surviving events are passed on to the drifting RFI filter.

Signals that drift rapidly in frequency and are thus are not detected by the passband RFI filter are detected and rejected by the drifting RFI filter. Event density in frequency time space is calculated for each dataset. A sectored semi-ellipse around each detection (see Fig. 2) is analyzed for sectors containing an excess of detections. Such sectors are scrutinized for signals that are coherent over multiple pointings of the telescope.

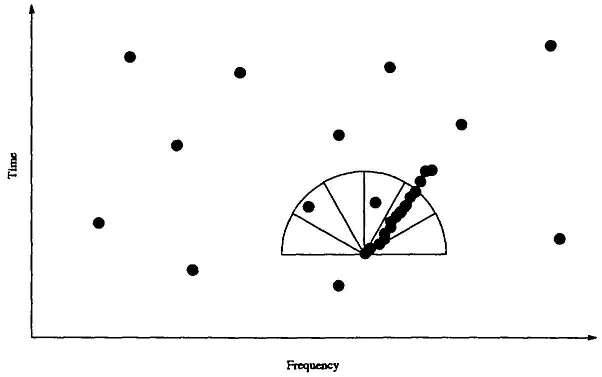


Figure 2. Illustration of SIREN's drifting RFI detection algorithm. A sectored semi-ellipse of frequency-time space around each detection is analyzed for signals that exhibit time-coherency and persistence over multiple pointings of the telescope beam.

5. Pattern Detection

The result of SERENDIP's suite of pattern detection algorithms is a collection of statistically interesting events. These events are correlated and merged to form an ETI candidate database. Algorithms are employed to detect signal persistence, telescope beam pattern matching, pulsed signals, and very high power events.

The primary pattern that we look for is persistence in frequency over time and position. All data are frequency corrected to the solar system barycenter. Thus, only those signals that display the proper Doppler drift relative to the solar system barycenter will remain at a nearly constant frequency. We employ two definitions of "persistence in frequency" in order to find two different classes of signals. The first class are those signals for which the Doppler effect of transmitter motion has been subtracted out. Signals in this class would show very little frequency drift and characterize a deliberately beamed message. This class is defined as containing signal sets that show no more than a 20 Hz frequency spread. This frequency window was arrived at by calculating the worst case frequency offset due to inherent pointing uncertainty; a strong signal could be detected a beamwidth away from its true source. One or more detections during each of two or more observations of a given sky position is logged as a candidate of this class. Radio leakage from another civilization will likely drift significantly over time spans of minutes to months. These significant drifts define the second class of signals for which SERENDIP is sensitive. Here, signal sets can have a frequency spread of up to 50 KHz. This window was arrived at as the worst case scenario of a transmitter moving with Jupiter-like acceleration. Two or more detections during each of two or more observations of a given sky position produce a hit of this class. There is also the additional constraint here that the same-observation detections must either not vary more than 20 Hz in frequency or display a coherent drift rate over multiple spectra.

Consecutive detections are analyzed for power distributions matching the telescope's gaussian beam pattern. This test is applied whenever we have 3 or more detections in consecutive integrations that are within 20 Hz of one another. A high degree of confidence in a beam pattern match indicates that the signal could be a point source on the sky.

Signals that are within 20 Hz of one another (or have a coherent drift rate) and occur at regularly spaced intervals and are cataloged as possible pulsed signals. Signals of particularly high power that do not persist over more beams than a point source would allow are also tagged as interesting.

6. Candidate Merging and Ranking

A SERENDIP candidate is defined as a one beamwidth area on the sky that has been identified by one or more pattern detection algorithms. A candidate is fully described by one or more candidate records (candidate merging), each of which is the result of detection by a single algorithm. In addition, celestial objects such nearby stars, globular clusters, and known planetary systems are entered into the candidate database to check for coincidence with SERENDIP candidates.

Candidates are ranked by two methods. First, each candidate record has a score, which is essentially the probability of this detection occurring by chance. Second, we ask how many algorithms detected the candidate. Here we consider coincidence with an interesting celestial object to be an "algorithm".

To determine the algorithm specific score, we calculate the relative probability of our detections occurring from noise, given the number of detections in the candidate area, the number of times we have observed the candidate area, the number of frequency bins assumed by the algorithm, and the actual frequency separation of the detections. The relative probability is given by

$$P = \frac{n_e^{n_d}}{n_d! \left(\frac{F_{tot}}{F_{win}}\right)^{(n_d-1)}} \left[\frac{\Delta f + f_\sigma}{f_\sigma}\right], \tag{1}$$

where: n_e is the total number of events logged at any frequency in the candidate area; n_d is the number of detections within f_{win}; f_{win} is the frequency window assumed by the algorithm; f_{tot} is the total SERENDIP band observed; Δf is the maximum frequency separation for the detection set; f_σ is the expected frequency variance.

If the candidate is assumed to be a signal for which the reference frame of the source has been corrected for, the dominant source of f_σ is position uncertainty.

If the candidate is assumed to not be source corrected, the dominant source of f_σ is Doppler drift induced by the motion of the transmitter.

The assessment of joint probability in the case of detection by more than one algorithm is more problematic. The pattern detection algorithms are not orthogonal to one another, thus the probabilities do not simply multiply. The ranking of multiple detection candidates is subjective but nonetheless useful in generating a list for re-observation.

The top low-probability and multiple-algorithm candidates are subjected to detailed check for RFI contamination and those surviving constitute the list of interesting candidates to be the subject of re-observation.

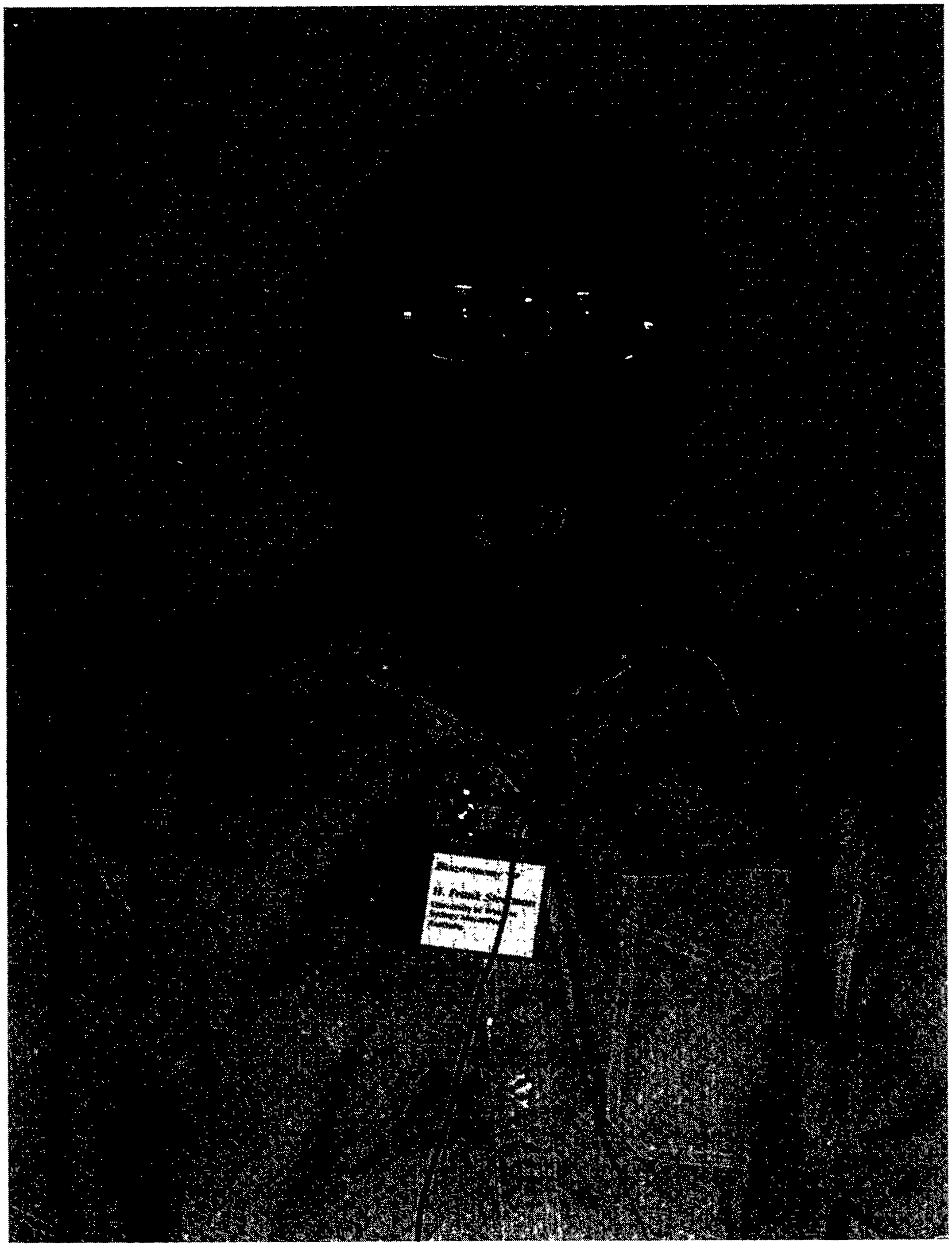
7. Candidate Reobservation

SERENDIP employs 2 types of candidate re-observation. The first type is simply chance re-observation during the normal sky survey. This typically happens every six months. As candidate beams are re-observed, additional detections or the lack thereof will alternately add or subtract from a candidate's score. In addition, SERENDIP has been awarded twenty-four hours of dedicated lookback time. This time will be used to re-observe the top scoring candidates. During dedicated lookback, sensitive narrow band on and off source observations will be used to characterize the candidate.

Acknowledgments. The SERENDIP project is jointly funded by the Planetary Society and the SETI Institute. The authors would also like to acknowledge generous donations from Mrs. Elizabeth Bowyer, Sun Microsystems, Toshiba, Intel Corp., Xilinx, and the fund raising efforts of the Friends of SERENDIP headed by Arthur C. Clarke. We also acknowledge science and engineering input from Michael Davis of Arecibo. SERENDIP's successful efforts are also dependent on the excellent support from lead telescope operator Rey Valez and the entire Arecibo Staff.

References

Werthimer, D., Bowyer, S., Cobb, J., Lebofsky, M., & Lampton, M. 2000, this volume



Frank H. Stootman