Auto scan—Technique for scanning masses of data to determine potential areas for detailed analysis

by DAVID L. SHIPMAN and CLARENCE R. FULMER

National Aeronautics and Space Administration
Huntsville, Alabama

INTRODUCTION

The Skylab Program is expected to provide advanced technology to assist in the development of large, permanent space stations. Skylab consists of four modules; the Orbital Workshop (OWS) which provides crew quarters and commodities, the Airlock Module (AM) which provides power distribution, atmospheric conditioning and services for Extra Vehicular Activity (EVA), the Multiple Docking Adapter (MDA) which provides for Command and Service Module (CSM) docking, and the Apollo Telescope Mount (ATM) which is essentially a manned solar observatory. Skylab will carry on board a number of experiments in scientific, technological, engineering, and medical areas including those in the ATM. The ATM is also an experiment. Several types of data will be gathered from the experiments including film, samples, observations, as well as transducer and thermocouple measurements. There will also be data from Skylab subsystems measurements since several subsystems have never been flown. The total number of measurements from the experiments and the Skylab subsystems will exceed two thousand measurements, a large number of which will be operating for twenty-four hours per day over the entire mission of eight months.

One million pages of computer printout for each twenty-four hour period may well be required to present all the data gathered from the measurements. The analysis of this volume of data either requires an excessively large number of engineers for the allotted twenty-four hour period or the development of some method for automatically selecting only those data which require some indepth analysis. The AUTO SCAN program is a computer program which searches all incoming data for data points which exceed some predetermined limit. The data points which are outside the limits are sent to the system analysts for detailed analysis. The system analyst is then relieved of the requirement to review all data and can concentrate on analyzing the exceptional cases. This corresponds to the management by exception principle.

DATA TRANSMISSION

The data from the Skylab measurements will be transmitted to the ground receiver station via telemetry at predetermined intervals. The receiving ground station will remove all redundant data points using a zero order algorithm according to a predetermined priority scheme and will transmit the resulting compressed data stream to the using NASA Center via data lines where it will be stored on magnetic tapes. The using NASA Center will then remove overlapping ground station coverage, chronologically order the data, and insert data from onboard recorders into the data stream where there are gaps in ground station coverage: the final data stream will then be put into a specific format called All Digital Data Tape or ADDT. One ADDT will be made for the AM telemetry system and one for the ATM telemetry system. These ADDT will become the primary input variable for the AUTO SCAN program.

PROGRAM CONSTRAINTS

Several major constraints and/or groundrules were baselined for the AUTO SCAN program as the first step in the development process. These constraints were:

1. The program must accept the ADDT and execute as the ADDT becomes available.
2. The program must be written in FORTRAN IV for use on the UNIVAC 1108 EXEC VIII computer using a maximum of 65K word core storage. Other languages are permissible only when it can be demonstrated that, for certain segments, another language is more desirable.
3. The program must be modular, such that modules can be removed or inserted without affecting the operational capability of the program.
4. The program is intended for use on the Skylab Workshop TM data to identify anomalous data and should not be designed to perform analysis or evaluation.
5. The program must be able to identify out of limit data and confirm the occurrence of specific events.
6. The computer run time must not exceed two hours for each set of data transmission which will occur about every six hours.
PROGRAM REQUIREMENTS

The next step in the development of AUTO SCAN was the establishment of the requirements for the program. The engineers responsible for each of the systems and subsystems as well as the experiments were expected to be the potential users of the AUTO SCAN program. Requirements submitted by these users were synthesized to identify the following types of information needed for the program development.

1. The measurement title and description and the associated processing priority required.
2. Related measurements that may be correlated with the desired measurement should be out of tolerance condition develop.
3. Measurement characteristics such as discrete/event, steady state or other (i.e., slopes, cyclic, consumables, etc.).
4. The time span over which an AUTO SCAN run is desired.
5. Special calculations that may be required before an AUTO SCAN run can be made.
6. The nominal value and the upper and lower test limit for each measurement. Changes in either of the limits that can be predetermined should be specified.

In reviewing user requirements, it was noted that many measurements were requested by several different users whose purpose for the measurement were completely different. As a result, the requests for measurement were completely different. As a result, the requests for measurement scans was significantly higher than the requests for measurements (on the order of about 1.5). This additional scanning capability compounds the core storage and computer run time problems. Accordingly, considerable coordination with the users was required to fit the requirements within the current AUTO SCAN program constraints. Ideally, the AUTO SCAN program’s physical development should have begun only after virtually all the requirements had been generated or at least until after a wide range of requirements had been received. Both Skylab development problems and manpower limitations precluded this ideal approach. An iterative approach was thus adopted whereby the overall planning was made for all anticipated requirements, but programming was done only on parts of a minimum capability program and as these parts were developed, additional capability was added.

PHYSICAL DESCRIPTION OF THE SYSTEM

A pictorial view of the AUTO SCAN program is presented in Figure 1. After several iterations, a baseline AUTO SCAN program was developed. This baseline program actually consists of four large subprograms.

1. **Input Processor**—The input for this subprogram is either cards or magnetic tapes reflecting the various requirements which are received. This subprogram arranges the requirements by system and provides the necessary scan instructions to the other subprograms.
2. **ATM AUTO SCAN**—This subprogram scans the ATM ADDT using instructions from the Input Processor and creates magnetic tapes containing the out-of-limit violations, the keying required to obtain plots or tabulations from another program during the out-of-limit interval and the data which may be required for some statistical calculations. Storage was allocated for 2047 scans of the ATM ADDT.
3. **AM AUTO SCAN**—This subprogram is essentially the same as the ATM AUTO SCAN except that the operations are performed on the AM ADDT.
4. **Output Processor**—This subprogram operates on the data generated in the two scan subprograms to present usable hardcopy output or printout.

The input to and output from the subprograms of the baseline AUTO SCAN program are shown in Figure 2. Some of the terms or acronyms used in this figure are defined in the Appendix.

Flow through these subprograms is expected to be as follows. The path is first through the solid lines and then through the dashed lines.

A generalized flow diagram of this baseline program is shown in Figure 3. This baseline program, with required storage for the telemetered measurements, uses about 50K word core storage for the UNIVAC 1108 computer. A number of Special Computation Modules are being developed and will be added to the baseline program as core storage and run time permits. The baseline AUTO SCAN
program is essentially input/output bound, i.e., the clock time is much larger than the CPU time. The Special Computation Modules will utilize the multiprogramming capability of the UNIVAC 1108 resulting in co-processing of the Special Computation Modules and the base line AUTO SCAN program, thus providing for more efficient utilization of CPU time. In addition, use will be made of mass storage devices to reduce requirement for computer core storage.

**PROGRAM OUTPUT**

There will be two basic outputs from the AUTO SCAN program: discrete events and out-of-limits intervals (flags). All the hard-copy output of the program will be in the form of tabulations. The information provided on each discrete event occurrence is as follows:

- **Measurement Number**
- **Measurement Title**
- **Time of Detection**
- **True or False Indication**

The information provided on each out-of-limits interval is as follows:

- **Measurement Number**
- **Measurement Title**
- **Time of Detection** — This is different from the start of the out-of-limit interval in that a persistency factor is applied to minimize the effect of noise.
- **Time of Interval Start** — When measurement initially goes out-of-limits.
- **Time of Interval Stop** — When measurement returns within limits.
- **Upper Test Limit**
- **Lower Test Limit**
- **Number of Data Points Outside the Limits**
- **The Maximum Excursion from the Limits**
- **The Average Value While Outside the Limits**

A sample output of the AUTO SCAN program for discrete is shown in Figure 4 and for flags is shown in Figure 5. On this particular test run, there were more than three times as many discretes detected as flags raised. The data used for the test run was a magnetic tape generated during the quality assurance checkout tests for the ATM only. Even with this enormous reduction in volume of data (approximately 10³) to be analyzed, additional effort is required to further reduce the volume of data to manageable proportions.

**ADDITIONAL DEVELOPMENT NEEDS**

Research in two areas could lead to additional reduction in volume of data to be analyzed: (1) Assessment of filtering feasibility and streamlining the AUTO SCAN modules and (2) development of onboard data redundancy removal and scanning. This second item merits some additional discussion.

Only meaningful data should be transmitted from the spacecraft to the ground station to reduce the load on...
both these systems and amount of data that would have to be processed by a ground based program such as AUTO SCAN. One way to accomplish this is by the use of algorithms. These algorithms should be developed for onboard data redundancy removal and should automatically scan the spacecraft system parameters as a minimum prior to transmission to the ground site. Criteria for removal and scanning algorithms are:

1. Redundancy removal techniques should have the capability of eliminating both totally redundant and near redundant data from the data stream.
2. Practical pattern recognition schemes should provide for the storage and transmission of the basic data characteristics rather than the total data stream.
3. Automatic scanning techniques should be capable of detecting data that falls outside a prespecified corridor and transmitting only this data.

These particular algorithms would apply basically to the onboard computer and obviously if implemented would perhaps impact both its design and capacity.

CONCLUSION

The ATM Baseline AUTO SCAN program has been developed and has received considerable testing during the August-January 1973 time period. Several Special Computation Modules are currently being developed and will be added to this baseline program to provide full capability for this Skylab module. Both the ATM AUTO SCAN and the AM AUTO SCAN subprograms have been completed, including the necessary checkout. The two subprograms have been used during Skylab ground checkouts and will be updated based on the results of its operation during these tests.

The final test of the AUTO SCAN program will occur during the mission. The program will be refined during the mission as experience is gained and will become a primary system analysis tool for future missions. The developers of the system are enthusiastic about the capability of AUTO SCAN and are already considering its use to handle the masses of data created in civil sectors as in the medical and environmental fields.

APPENDIX

Requests—Requirements submitted by the potential users of the AUTO SCAN program which are punched onto cards and subsequently onto magnetic tapes.

Calibration Data—Measurement calibration tape which is also transmitted via data lines from other Centers and used to convert data to engineering units.

Compact Tape—Magnetic tape consisting of the requirements and specific instructions to be used in subsequent subprograms.

ADDT—All Digital Data Tape—Telemetered measurements reformatted, chronologically ordered with redundant ground station coverage removed and onboard tape recorder data inserted where data gaps occur.

Special ADDT—ADDT created by another major program which can be directly input to the AUTO SCAN program without the necessity of using an ADDT read routine which saves core storage and computer time.

Behavior History Tape—Magnetic tape containing the information relative to the time interval when the measurement exceeded the out-of-limit tolerances.

BHT Keying—Output which indicates to subsequent programs that either correlated or other measurement data either in tabulation or plots are required during the out-of-limit intervals provided on the BHT.

Statistical Data Tape—A magnetic tape that will contain data required in the event that some statistical calculations are needed.