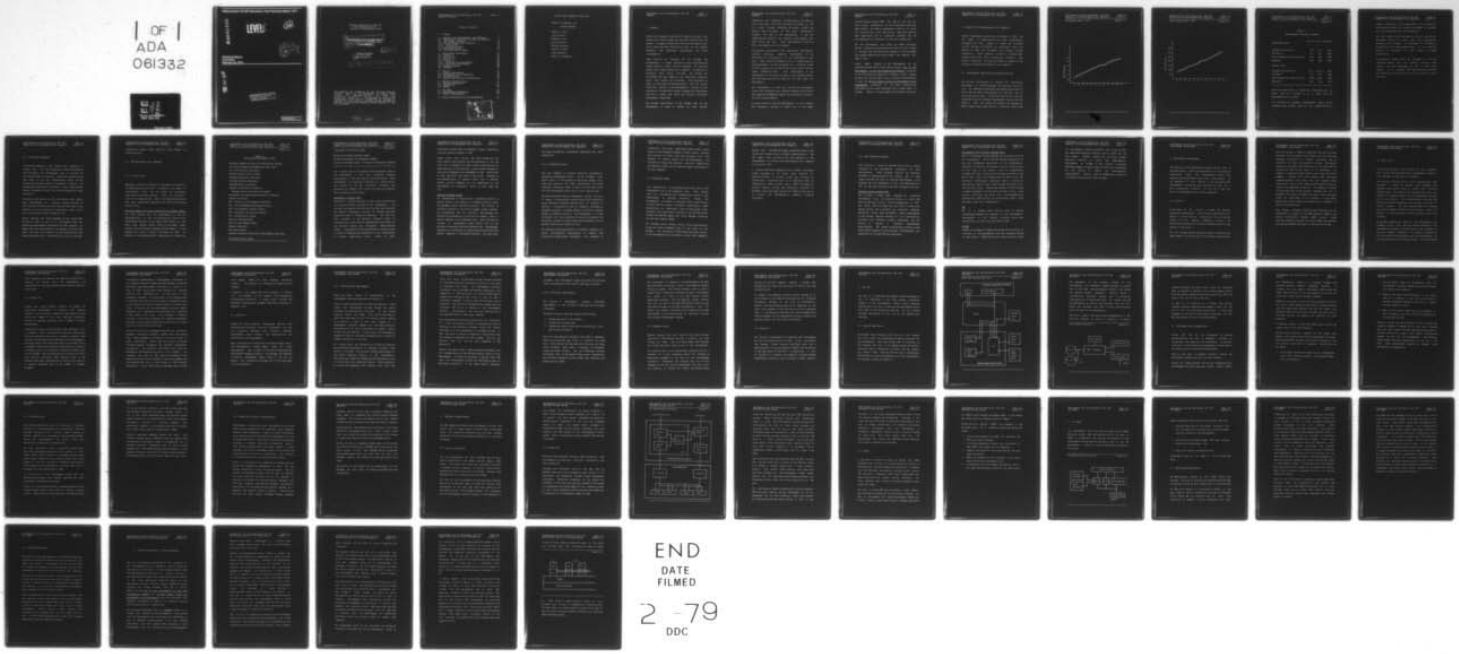
AD-A061 332 COMPUTER CORP OF AMERICA CAMBRIDGE MASS F/G 17/2
DATACOMPUTER AND SIP OPERATIONS.(U)
FEB 78
UNCLASSIFIED CCA-78-04 MDA903-77-C-0156 NL

| OF |
ADA
061332



END
DATE
FILMED
2-79
DDC

Datacomputer and SIP Operations: Final Technical Report, 1977

LEVEL II

Technical Report
CCA-78-04
February 24, 1978

Computer Corporation of America
575 Technology Square
Cambridge, Massachusetts 02139

Computer Corporation of America
575 Technology Square
Cambridge, Massachusetts 02139

6 Datacomputer and SIP Operations

9 Final Technical Report, 1977

11 24 February 1978

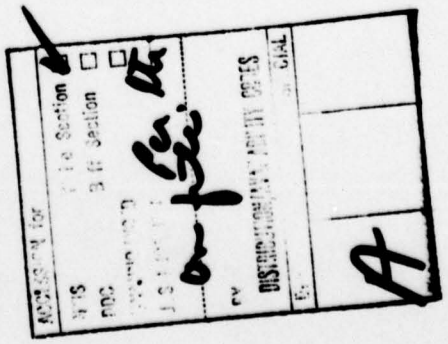
12 60p.

Technical Report
14 CCA-78-04

15 MDA903-77-C-0156

This research was supported by the Advanced Research Projects Agency of the Department of Defense under Contract No. MDA-903-77-C-0156. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

Table of Contents



Project Staff Members at Year End

Donald E. Eastlake, III

Project Manager

Robert H. Dorin

Jerry Farrell

Stephen A. Fox

David Kramlich

Matthew Maltzman

Gerald E. Maple

James Schmolze

Steven A. Zimmerman

1. Summary

During 1977 Computer Corporation of America provided very large on-line data storage and retrieval services over the Arpanet to support seismic data activity and general use. Use of these services continues to grow in the Arpanet community and additional applications are under investigation.

These services are provided by CCA through the Datacomputer, a system designed to allow convenient and timely access to large on-line databases for multiple remote users communicating over a network. In addition to operating these direct services, CCA assists and coordinates the user community and additional potential users, both seismic and non-seismic. This assistance is aimed at achieving the maximum benefit from the unique facilities offered by the Datacomputer. As part of this assistance, CCA maintains several programs and subroutines that run on remote user hosts and provide convenient Datacomputer interfaces.

The seismic application is the largest user of the Datacomputer in terms of amount of data stored,

complexity, and bandwidth. It sends much of its data to CCA in real time. This real time data is fielded at CCA by a small reliable dedicated processor, called the Seismic Input Processor, or SIP, which periodically forwards the data to the Datacomputer. A new SIP communications protocol and numerous improvements were made during the year. Final improvements to meet all known requirements are now complete.

Evolutionary improvement in the operational Datacomputer software continues. Numerous enhancements to the efficiency and accessibility of the Datacomputer were made. Some active development work on a system based on the Datacomputer is being conducted by a different group at CCA in support of the ARPA ACCAT project under Contract number N00039-77-C-0074. Such improvements in the Datacomputer as have been included in this separate effort have been made available to users of the TBM based CCA Datacomputer.

The enhancements to both the SIP and the Datacomputer during 1977 stretched their software capacity sufficiently that swapping debugging programs were installed in each to free more program memory.

A unique feature of the CCA Datacomputer is the copious and inexpensive storage it offers due to its Ampex

Tera-Bit Memory System (TBM). The TBM at CCA was the first public installation of this video-tape technology based system. Our TBM is configured to hold up to about 175 billion bits on four tape drives. Some down time has been experienced due to controller problems and the non-redundancy of controllers in the CCA installation.

The CCA Datacomputer runs under our TENEX operating system. Extensive modifications were made to this system in the past and some additional changes to increase TENEX file system capacity and handle added TBM features were made in 1977.

Formal papers related to the Datacomputer and SIP presented during 1977 include the following: Use of the Datacomputer in the Vela Seismological Network, presented at the International Symposium on Computer Aided Seismic Analysis and Discrimination held in Falmouth Massachusetts in June, and Tertiary Memory Access and Performance in the Datacomputer, presented at the Third International Conference on Very Large Databases held in Tokyo Japan in October. Copies of these papers are available from CCA.

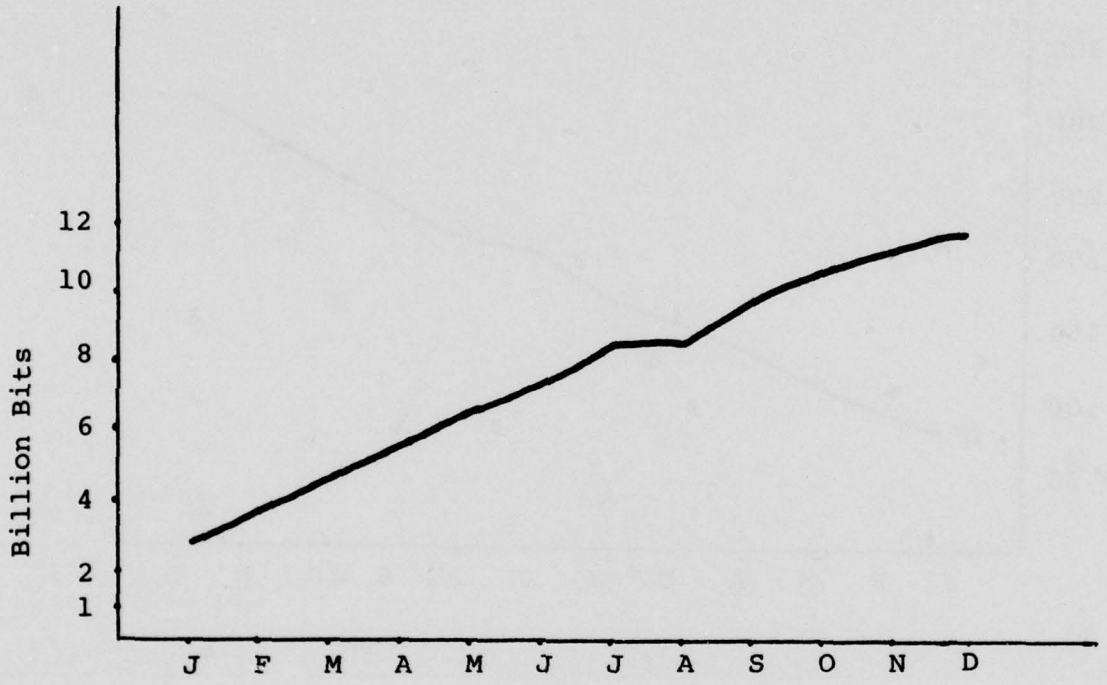
2. Support of the Datacomputer User Community

Regular Datacomputer service was introduced in 1977, and the number of applications has grown substantially throughout the year. Figure 2.1 shows the increased on-line storage utilization by non-seismic users and Figure 2.2 shows the increased total seismic storage utilization. Other measures in Table 2.1 indicate that system activity has shown significant increases. This section summarizes the tasks performed in support of the continually growing Datacomputer user community.

2.1 Datacomputer Operations and General Services

CCA strives continuously to improve the operational performance of the Datacomputer from the user's point of view. All scheduled preventative maintenance was moved to before 9AM and 9AM to 7PM weekdays was originally set as our prime service time. During this time we will take all reasonable steps to maintain Datacomputer service over the network. Later the prime time interval was extended to 8PM to improve west coast service. Finally an 11PM to 7AM

Non-Seismic Data - 1977 Figure 2.1



Total Seismic Data - 1977 Figure 2.2

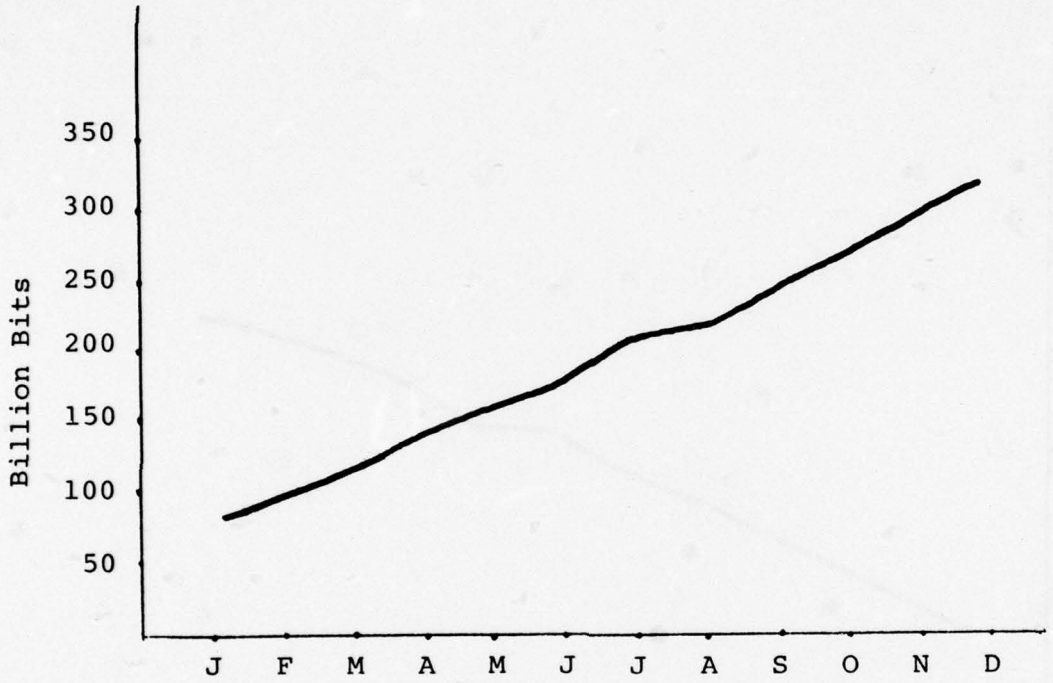


TABLE 2.1

Datacomputer Utilization Summary

	Jan 77	Dec 77	increase
Non-Seismic Users			
Storage (billion bits)	2.8	11.6	314%
CPU Hours	10.1	40.3	299%
Data transferred (billion bits)	436	1346	209%
Sessions	1739	3995	130%
Seismic Users			
Storage (billion bits)	80.0	324.9	306%
CPU Hours	41.8	78.9	89%
Data transferred (billion bits)	8456	19293	128%
Sessions	478	702	47%

shift was added which is generally unattended but for which users are informed if it is known that the Datacomputer will be down.

CCA provides an automatic Datacomputer status server program which advises users of the system status and

network connection to the Datacomputer, this program has been enhanced recently to indicate the number of messages and bits transferred over each connection.

The user support staff continued to provide a 24-hour per day "HELP" service to assist users with Datacomputer questions and problems. A Programmer of the Week is available at CCA during working hours, and an answering service is available at other times for urgent difficulties.

An accounting system which was designed in 1976 was completed during 1977, and, recently, detailed usage reports have been generated and sent to users upon request. It is intended that these accounting reports will eventually form the basis of direct billing for costs incurred.

2.2 The Seismic Community

CCA provides support to the seismic user community in several ways. We have frequently reviewed Datacomputer file structures and datalanguage used in storing and retrieving seismic information. Particular assistance has been given to the Vela Seismological Center and the Seismic Data Analysis Center in Alexandria Virginia, the Albuquerque Seismological Laboratory, Lincoln Laboratories Applied Seismology Group, and Texas Instruments (an SDAC contractor).

During the last quarter of 1977, CCA helped LL-ASG modify their datalanguage for a complex seismic retrieval of short period non-array data and consulted with SDAC on the effects on on-line retention of terminating direct short period array data storage through the SIP.

Donald Eastlake and Steve Zimmerman of CCA visited SDAC and VSC on June 14 and 15 to investigate means for maximizing Datacomputer performance for SDAC requests. On August 26th, CCA participated in a meeting at Lincoln with ARPA, VSC, SDAC, and LL-ASG to lay the foundation for the Seismic Waveform File. The SWF will be a file of the most

significant Seismic data, primarily that related to detected events.

2.3 The Non-Seismic User Community

2.3.1 Active Users

Aggregate statistics indicate an increased utilization of the Datacomputer during 1977. In addition, the number of organizations which use the system for non-seismic applications has grown steadily. Table 2.2 presents a list of the organizations. Some of the applications which have made significant progress during 1977 are described briefly below.

Advanced Command and Control Architectural Testbed (ACCAT)

The ACCAT program uses the Datacomputer at CCA for experimentation, as well as operating two copies of the Datacomputer software in the secure ACCAT network at the Naval Ocean Systems Center (NOSC) in San Diego. These systems provide the basic database system support to the command and control activity performed at NOSC. In addition to the assistance and consultation provided by

TABLE 2.2
Non-Seismic Datacomputer Users

Advanced Command and Control Architectural Testbed

Air Force Armament Development and Test Center

Air Force Avionics Laboratory

Arpanet Network Control Center

Brookhaven National Laboratory

Lawrence Berkeley Laboratory

Carnegie-Mellon University

Computer Corporation of America -

Message Archive and Retrieval System

Department of Energy/Argonne National Laboratory

Harvard University

MIT - Artificial Intelligence Laboratory

MIT - LCS Programming Technology Division

MIT - LCS Database Systems Division

MIT - Laboratory for Nuclear Science

MIT - Plasma Fusion Center

National Software Works

Rome Air Development Center

Rutgers University

SRI International

Stanford University Artificial Intelligence Laboratory

(continued on next page)

Table 2.2

(continued from previous page)

Stanford University Medical Center

US Army Development and Readiness Command

University College London - Facsimile Transmission Research

CCA in areas such as file design and performance, many of the enhancements to the user interface software incorporated in 1977 were done for the ACCAT users. These enhancements are described in Section 2.5. Specifically, the creation of the TAP relational interface and extensions of DCSUBR to the Fortran and Cobol environments were requested by ACCAT users.

Department of Energy (DOE)

Substantial progress was made in 1977 toward the creation of a national weather database on the Datacomputer. CCA actively coordinated this activity, and has been assisting the Applied Mathematics Division at Argonne National Laboratory (ANL) in loading the database. The data has been provided by the National Climactic Center in conjunction with the National Oceanographic Data Center of the National Oceanic and Atmospheric Administration (NOAA). Experiments are being performed by groups at ANL in order to determine the suitability of the Datacomputer in several application areas. Groups at other

laboratories within DOE are expected to begin experiments with the weather database in 1978.

Robert Dorin, Jerry Farrell and David Kramlich of CCA spent several days working with DOE personnel at the ANL site, and, on December 1st, gave several presentations and demonstrations of the partially loaded weather database to applications managers and programmers at ANL. Ronald Bare of ANL describes the effort to this point. "Because of this demonstration, it appears likely that a national weather history database will be established on the Datacomputer as a cooperative effort by DOE, NOAA and CCA."

Facsimile Message System

The Datacomputer is being used as a database system for a project aimed at the creation of a message system in which the messages contain facsimile pictures. This effort is currently underway at University College, London, England. The Datacomputer and its interface, Datalanguage, are particularly appropriate in this sort of system, where a distant computer is the immediate user. There has been an increase in Datacomputer usage during 1977 by the facsimile researchers who have reported that Datalanguage responses are "excellent for synchronizing activities with another computer or intelligent terminal. They give just

the right information on successful compilation and error conditions."

2.3.2 Prospective Users

CCA also attempts to provide technical assistance to prospective Datacomputer users. During December, 1977, Robert Dorin of CCA visited an Army group, ALSMA, in St. Louis and a group in San Diego representing the Naval Underseas Technology Center of NOSC to advise them on the role the Datacomputer might take in their applications.

The Army Automated Logistics Management System Activity, or ALMSA, is developing a system called ELITE (Executive Level Interactive Terminal System) which is intended to provide a collection of tools for managers, including an automated calendar, a message system, a milestone tracking system, and keyword archives. The Datacomputer is being considered for this collection to provide economical and convenient storage of large volume data components such as keyword archives and historical data.

The Underseas Technology Center is creating a database of marine environmental measurements to support the activities of NOSC marine biologists. The database is

potentially very large. Researchers would access a small subset of the database relevant to their current interest and retrieve it to their local site (a Univac 1110). This scenario is ideally suited for the Datacomputer. Its inexpensive storage and extensive selection capabilities are employed along with the superior numeric processing of the host computer.

2.4 Background Usage

Four applications on the Arpanet which store data in the Datacomputer run in the background mode. These systems which run continuously will automatically connect to the Datacomputer at periodic intervals. Though the Datacomputer is subject to periods of heavy load and must be taken down for regular preventive maintenance, these applications are unaffected. They are the Seismic Input Processor (SIP) described in Section 4; the MIT-DMS Survey System; the BBN IMP Logger; and CCA's Message Archiving and Retrieval System (MARS).

The MIT-DMS Survey system stores files containing the status and connect response time of the hosts on the Arpanet. The quantity of Survey data being kept on-line in the Datacomputer rose from 600 to nearly 1200 megabits

during 1977. The BBN IMP Logger accumulates data on IMP status and irregularities in Arpanet communications. The IMP Logger files currently use 828 megabits of the Datacomputer, an increase from approximately 600 megabits at the end of 1976.

A recently developed background user is MARS, a prototype system developed by CCA (under ARPA Contract No. N00014-76-C-0091) to provide filing and retrieval of Arpanet mail. Electronic mail is an important mode of communication among Arpanet users. MARS is intended to provide economic storage and convenient retrieval of messages which have been stored into complex files set up to utilize the Datacomputer's automatic indexing facilities.

2.5 User Interface Software

CCA provides a series of programs which offers a simple interface to the Datacomputer convenient for many applications. These programs simplify the technical problems of communicating on the Arpanet, and, in many cases, preclude the need for learning Datalanguage. A new program, TAP, was implemented in 1977, and enhancements to TAP and the other programs continued throughout the year.

Terminal Access Program (TAP)

TAP provides a simple query language for accessing Datacomputer files in relational format, i.e. flat files with no duplicate records. TAP was implemented in 1977, and allows users to create and load files, to select and update records, and to perform the relational algebraic operations called JOIN and PROJECT. Recent enhancements to TAP include improved editing of file descriptions, operation from different network interfaces, and the ability to connect to different Datacomputer installations. The latter is particularly useful in the ACCAT secure network in which multiple Datacomputers are accessible for reliability and efficiency.

Datacomputer File Transfer Program (DFTP)

DFTP is a terminal-oriented package for archiving files on the Datacomputer. Use of the DFTP system continued to be the most widespread application of the Datacomputer. Currently 22 sites use the program to store 8 billion bits of file data on the system. This is a significant increase from the 1.3 billion bits stored with DFTP at the end of 1976. One of the largest users of DFTP is the MIT Artificial Intelligence Laboratory where a variety of applications require large volume storage available at low cost and short delay. Recent improvements to DFTP include a command (EXAMINE) allowing direct perusal of text files without retrieving the whole file, and optimized file descriptions, MANY and BIG, for storing many small files and fewer large ones, respectively.

RDC

RDC is a program which assists users in sending Datalanguage commands and requests to the Datacomputer. Improvements to the terminal interface which were requested by its users have been made to RDC.

DCSUBR

DCSUBR is a package of TENEX subroutines which provide an interface to the Datacomputer from user programs running on remote hosts. DCSUBR serves as a basic building block

for both general interface software, such as TAP and RDC, and application-specific software, such as MARS and the IMP Logger. DCSUBR handles the low-level network protocols required in using the Datacomputer. Improvements to DCSUBR in 1977 include a modified buffering scheme to reduce network traffic and turnarounds and the ability to handle two Datacomputers simultaneously. Special forms of DCSUBR were created to be callable from Fortran, Cobol and BCPL.

3. Datacomputer Development

The progress of the Datacomputer software during 1977 is described below. Since the Datacomputer is in a primarily operational, rather than developmental phase, this progress was evolutionary rather than revolutionary. At the beginning of 1977, Version 3 was the operational Datacomputer. During the year, Versions 3/5, 3/6, and 4 were successively installed. At the end of 1977, the Version 5 Datacomputer was nearing completion.

3.1 Version 3

At the start of 1977, Version 3 became the standard operational Datacomputer. The principal improvements over Version 2 which Version 3 provides are the file groups feature and four special arithmetic function for determining distances and directions between points on the surface of the earth.

The file groups feature provides a means of handling the large number of files involved in the seismic application.

With file groups, a number of physical files with the same structure can be treated as one logical file for retrieval purposes. Furthermore, a "logical constraint" may be specified for each constituent of a group. The logical constraint indicates some restriction on the data which is asserted over the file for which it is specified. For example, a constraint might require that a date field fall between 1 January 1978 and 31 January 1978. Retrieval requests against a file group will only examine those constituent files which might contain qualified data in light of the specified constraint. For example, a request for data with date fields between 1 December 1977 and 31 December 1977 would not examine any data in the file whose dates are constrained to be in January. In the seismic application, data streams are divided into a sequence of daily or monthly files.

Four special arithmetic functions were developed for the Datacomputer in support of the ARPA Advanced Command and Control Architectural Testbed (ACCAT) project. These functions compute the great circle and rhumb line distance and bearing between two points on the earth's surface.

3.2 Version 3/5

In the first quarter of 1977, Version 3/5 was installed. This version had some operational improvements related to error messages and problem file flagging.

As a result of the evolution of more sophisticated user programs, a number of specific error messages were assigned unique prefix codes to replace their previous default codes. The simplest user programs just send a set sequence of computed requests to the Datacomputer and give up on any failure. More sophisticated programs take different branches depending on the success or failure of various requests. The most sophisticated user programs want error messages from which the particular reason for failure is easily parsible so that corrective action can be taken.

An operator command was added to the Datacomputer for flagging active files that are causing problems for the Datacomputer hardware or software due to the remnants of previous hardware problems. It is normally necessary to fix such problem files manually while the Datacomputer is not providing service to multiple users. To stop problems

from cascading, the operator can flag the problem file to prohibit all access until the Datacomputer can conveniently be shut down outside normal service times for repair work.

3.3 Version 3/6

During the second quarter, Version 3/6 became the operational Datacomputer. It features more general inverted file retrievals, doubling of the number of files that can be staged at one time, improved statistical records of staging activity, and a number of lesser improvements.

Inverted retrievals are particularly fast references into a database. Such retrievals use auxiliary index tables that are selected by the user at database creation time and then are automatically maintained by the Datacomputer. Previously these fast references could only be made if the information being retrieved was characterized in terms of a constant value or the value of a declared variable local to the request. This was generalized in Version 3/6 to do fast inverted retrievals where possible even if the desired items where characterized in terms of the value in a general container such as an element in another database.

An important consideration in Datacomputer efficiency is to minimize transfers between working storage on disk and the slower mass TBM storage. One way to do this is to keep as much active data on the staging (i.e., working) disks as will reasonably fit. In the past, it was sometimes necessary to discard data on the staging disks not because they were full but because of inadequate space in the active file table to describe what file pieces were there. In Version 3/6, the Datacomputer's virtual memory structure was overhauled to provide space for a doubling of the active file table. This produced a dramatic increase in the time an average file remained immediately available on the staging disks from on the order of a half hour to on the order of four hours.

Version 3/6 also has an expanded repertoire of statistics messages concerning transfers between TBM and staging disk. These messages are written to a log file which is saved indefinitely. These logs can be used to simulate different staging strategies and optimize Datacomputer performance. For transfers from TBM to disk, there are messages at the beginning and end of the transfer. For transfers from disk to TBM, there are messages at the beginning and end of the transfer of each data segment and the beginning and end of each clump of TBM hardware operations. In all cases there are messages when any TBM

error occurs. There are also messages concerning allocation and deletion of staging disk space and file opening.

A number of minor changes were also included in Version 3/6. For example, at the request of the Albuquerque Seismological Laboratory, a unique prefix (+U199) was assigned to Datacomputer messages indicating difficulty in opening a file.

3.4 Version 4

During the third quarter, Datacomputer Version 4 was developed and installed. This new Datacomputer was put into operation September 25th. Version 4 features improvements which enhance the accessibility and efficiency of the Datacomputer.

Some difficulty was encountered in finding enough room in the Datacomputer's virtual memory space for these improvements. This problem was solved by modifying the Datacomputer debugging program. The debugger was modified to swap the Datacomputer symbol table in and out of memory, thus avoiding the storage cost of this data when it is not required.

3.4.1 Accessibility Improvements

There are three levels of accessibility to the Datacomputer that were improved with Version 4.

First, initial accessibility by users was improved. Users access the Datacomputer through subjobs and subjob accessibility was improved in two ways. An idle subjob time-out feature was added. This aborts and logs off users that have been idle for five minutes without logging in or idle for a half an hour after logging in. A Datacomputer operator command was also added whereby a selected user's job can be eliminated in a manner similar to the idle job time-out. In both cases, the Datacomputer attempts to send an explanatory message before closing the network connections to the user.

At a second level, the accessibility of data was enhanced when preventive maintenance or hardware problems limit access to the TBM. This was done through improvements in the device handling routines in the Datacomputer. Facilities were added for suspending TBM operations on one or more drives in a controlled fashion. While operations on a drive are suspended, data recently read from that

drive will still be available on disk and data destined for that TBM drive can be written to disk up to the limits of available disk space. All Datacomputer directory information can be accessed and modified regardless of any suspension of TBM activities. If a user program makes a request that references data which is only on TBM, a distinctive message is sent to the user and their job is suspended. The job will be automatically resumed when the referenced TBM drive is enabled by the Datacomputer operator. Alternatively, the user can interrupt out of the suspended state to make other requests.

This suspension mechanism was chosen over the alternative of a new error return so that no modifications would be necessary to the automatic programs around the network that make us of the Datacomputer in a background mode. If data they wish to access is temporarily unavailable, they will simply be suspended and later resumed. The idle job time-out does not run for jobs suspended on TBM operations.

Finally, the third level of TBM accessing improvement was the implementation of the TBM Read Recover feature in the Datacomputer. If there are problems in reading from TBM tape, the Datacomputer invokes this TBM feature which is described in Section 5. If the Read Recover operation

succeeds, the Datacomputer takes care that the recovered data is ultimately written back on TBM tape correctly.

3.4.2 Efficiency Improvements

The Version 4 Datacomputer contains efficiency improvements in the writing of TBM tape and in request processing.

Reliable writing on TBM tape requires three steps:

1. erasing the block to be written;
2. writing the new data; and
3. reading the newly writen data to verify that it has been written correctly.

This is a slow operation because of the need to repeatedly start-up and stop the tape drive to traverse the same area. This process has been made more efficient for those (very frequent) cases where a sequence of contiguous blocks are being written. The new technique involves performing each of the steps (erase, write, read-verify) for the entire sequence at once rather than handling the blocks one at a time.

The processing of requests in the Datacomputer was made more efficient in two ways. First, a pre-compiled request feature was added. This feature enables a user to store a request under a name and to execute the same datalanguage any number of times in the same session without compilation overhead. This pre-compiled request feature was done as part of the ARPA IPTO-ACCAT project and was also made available on the TBM Datacomputer at CCA. The second way request efficiency was increased was through improvements in the routines used to interlock critical areas between Datacomputer subjobs.

3.5 Message Lockups

Network lockups have been found to occur under unusual conditions of Datacomputer use and a solution has been devised for them. They occur when simple user programs try to make particular types of use of the Datacomputer.

In a typical case, a user program pays attention only to a transfer on a data connection while the transfer is generating messages to the user on the Datalanguage connection. Eventually the pipeline may fill up with messages to the user and the Datacomputer will wait until the pipeline is cleared out before providing further

service to this user program. However, a simple user program will not take this action and so the job will hang up indefinitely.

An elegant and complete solution to these lockups is for user programs to use separate processes for the different information streams to and from the Datacomputer. However, multiple processes may not be supported on the user's host computer or the user may not wish to implement them. A satisfactory alternative has been designed which buffers messages at the Datacomputer during a transfer and supplies them to the user after the transfer.

3.6 Version 5

The Version 5 Datacomputer is presently under development and will be installed early in 1978. It will incorporate the message lockup solution described above and a generalization of the pre-compiled request feature mention in the description of Version 4. In particular, it will be possible to preserve pre-compiled requests between Datacomputer sessions as well as within a single session.

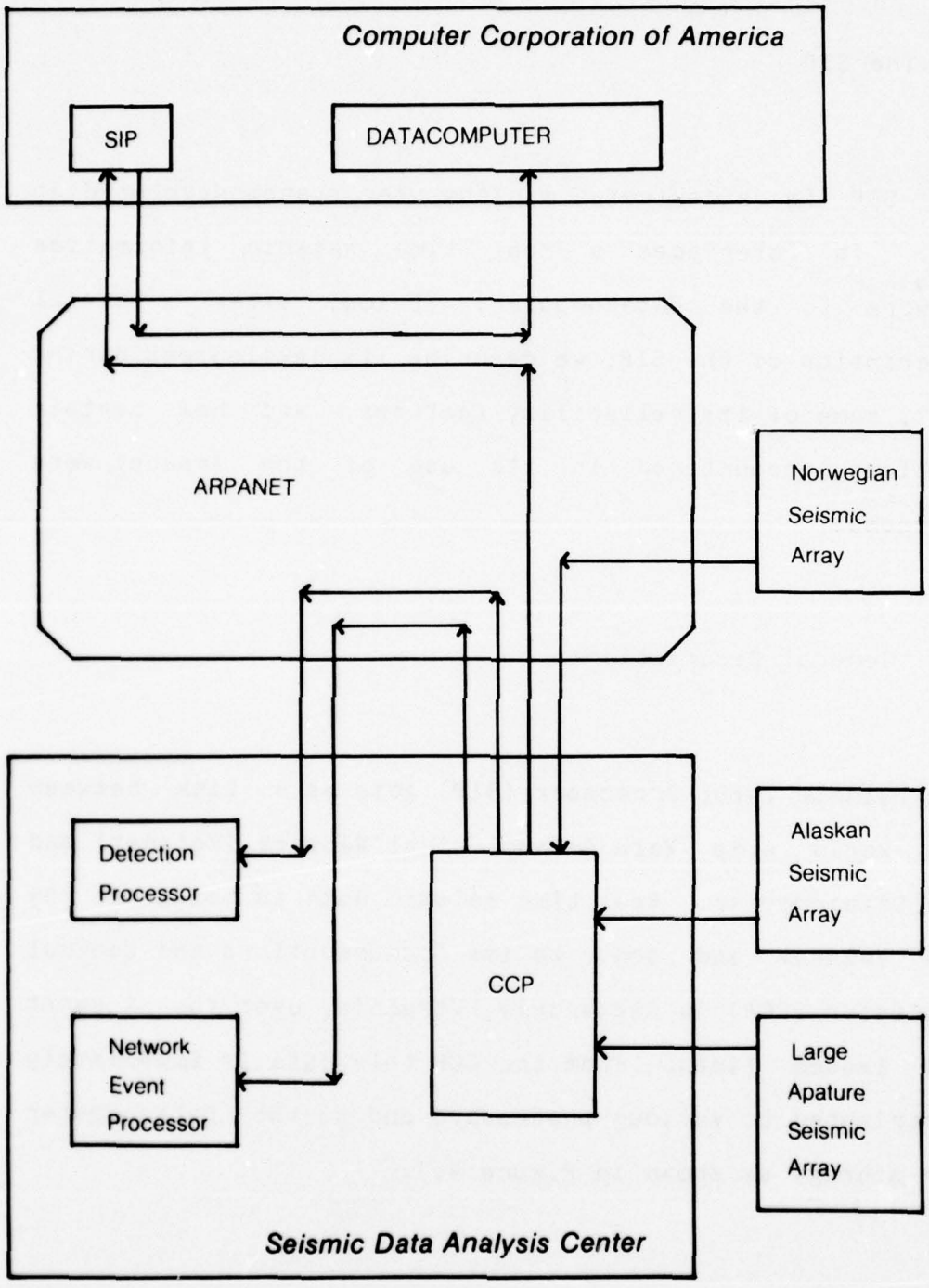
4. The SIP

The SIP is a dedicated minicomputer system developed at CCA. It interfaces a real time seismic information network to the Datacomputer. Below, after a general description of the SIP, we describe its development during 1977, some of its reliability features, and how certain problems encountered in its use of the Arpanet were overcome.

4.1 General Description

The Seismic Input Processor (SIP) acts as a link between the world wide Vela Seismological Network (Velanet) and the Datacomputer. Real time seismic data is collected by the Velanet and sent to its Communications and Control Processor (CCP) in Alexandria, Virginia, over the Arpanet and leased lines. From the CCP this data is immediately distributed to various processors and to the Datacomputer for storage as shown in Figure 4.1.

Real Time Seismic Data Flow Figure 4.1

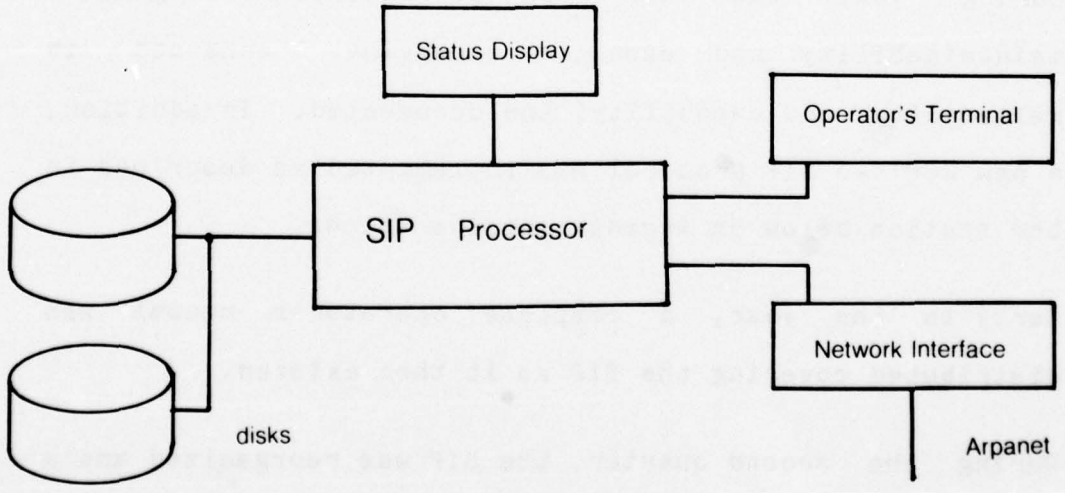


All components of the Velanet, except for the Datacomputer, are dedicated systems designed to receive data in real time. The Datacomputer, however, also serves the general Arpanet community, operates within a non-real-time operating system, and is periodicly unavailable due to maintenance work. As a result, the SIP was implemented to receive real time data from the CCP, buffer and reformat this data on disk, and periodicly burst the data to the Datacomputer.

The SIP is a small, dedicated system implemented on a DEC PDP-11/40 computer as shown in Figure 4.2. It has an

---

SIP Hardware Structure Figure 4.2



Arpanet interface, two RP-04 disk drives for buffering data, an operator's terminal, and a status display screen. With the present bandwidth and disk structuring, about two days of data can be held by the SIP.

As part of its function as a Velanet node, the SIP provides operator communications between itself and the CCP as well as processing seismic data. It also sends messages to the CCP for each chunk of data when the data has been properly filed in the Datacomputer.

4.2 Development and Documentation

During 1977 the SIP was reorganized for greater maintainability and ease in debugging, enhanced in reliability and capability, and documented. In addition, a new CCP <-> SIP protocol was implemented as described in the section below on Arpanet considerations.

Early in the year, a complete operator's manual was distributed covering the SIP as it then existed.

During the second quarter, the SIP was reorganized and a new debugger and basic operating system, called STRUGL,

was implemented. STRUGL is a swapping debugger that provides powerful symbolic debugging features. To minimize its core memory requirement, STRUGL keeps the symbol table and most of its code swapped out on disk and keeps only a small kernel normally resident.

STRUGL is also able to write and load programs to and from the SIP's disks, maintain directories of these programs, load from paper tape, downline load from CCA TENEX, copy selected portions of the disk on either SIP disk drive to the other, and set up default trap handling for user programs.

A detailed manual, called the "STRUGL User's Guide" was issued, also in the second quarter.

During the third quarter, updated SIP and STRUGL user manuals were issued after further enhancements were made. These enhancements included the disk pack changes and the new protocol described in subsequent sections and the following:

1. Up to eight seismic data sites can be accommodated with only directory changes. (No program changes will be necessary.)

2. The SIP avoids using the Datacomputer when the Datacomputer is heavily loaded unless the SIP disks are nearing their capacity.

3. Whenever the local CCA IMP indicates it is going down by a message to the SIP, the SIP sends a computed "operator" message to the CCP to inform the CCP operator of this.

4. When the SIP senses a power failure it attempts to send a message to the CCP and CCA IMP indicating it is going down and why.

Finally, during the fourth quarter of 1977, a manual was issued for NLDSIP, a program used to load other programs into the SIP system, and some improvements were made in the SIP's ability to recover from network difficulties between the SIP and the Datacomputer. Most previous network effort had been concentrated on the SIP - CCP path. These improvements required no changes in the user/operator documentation.

4.3 SIP Reliability

The SIP has achieved its goal of providing a reliable interface to the Datacomputer for Velanet data. The only problems encountered of a chronic nature related to Arpanet capacity as discussed in the following section. During 1977 approximately 200 billion bits were successfully transfered through the SIP.

The last algorithmic flaws in the SIP of which we are aware were eliminated during the first quarter of 1977. These former flaws related to a possible buffer lockup under unusual conditions and a queue overflow that could happen if the SIP received an extremely rapid burst of short Arpanet control messages as occurred once due to problems with the PLURIBUS IMP at SDAC.

In the process of its development during 1977, two mechanisms were added that greatly improved the SIP's operational reliability as follows:

First, an auto-reload facility was added whereby the SIP program reloads itself from disk for certain internal errors. When such errors are detected, an extensive type

out of the internal condition of the SIP is made including the complete contents of the status display screen. If set to the normal unattended mode, the SIP then reloads itself unless this is the second error within ten minutes, indicating a chronic or recuring problem. This auto-reload feature, in conjunction with the SIP's power fail restart logic, made it exceptionally robust.

Second, a disk pack changing feature was added. This involved adding code to frequently write out updated disk directories on each pack in case it is dismounted and to validate directory information read in from a newly mounted pack. Some additions were also necessary to the already extensive disk "error" recovery code to handle a drive being turned off and the pack removed at any point.

4.4 Arpanet and Protocol Considerations

Considerable difficulties were encountered early in the year using the Arpanet for the continuous high bandwidth seismic data transmission. Typical of these difficulties were indications from the network that messages could not be transmitted due to insufficient resources and excessive blocking of network output due to congestion. Early in 1977 it was determined that the prime causes of these problems were lack of reassembly space in the CCA IMP and inefficiencies in the CCP <-> SIP protocol. The solution to the difficulties was to eliminate these causes.

First, a new protocol for communications between the SIP and CCP was designed and implemented in early 1977 and became operational during the second quarter. This new protocol was designed by SDAC, CCA, and BBN to approach the most efficient possible use of the Arpanet. It achieves this primarily by packing logical messages into efficient, maximum size physical messages. The protocol previously in use transmitted each logical message as a separate and smaller physical message. Since the new protocol can split logical messages between physical

messages, there is a minor cost in software complexity and table space to reassemble the resulting logical message fragments. This is more than made up for by the reduced overhead within the network and improved network response.

(Simultaneously with adding the new protocol, the SIP was modified to accommodate data from a third seismic array site and to take into account the reduction in the volume of long period data arriving from the NORSAR site.)

Second, the lack of reassembly buffer space in the CCA IMP was solved by installing a PLURIBUS IMP at CCA during the third quarter of 1977. This PLURIBUS IMP was configured with enough buffer space for more than ten full size messages whereas the previous CCA IMP had room for only three.

The new CCP <-> SIP protocol and the installation of the PLURIBUS IMP have cured the network problems that were encountered.

5. TBM Mass Storage System

The TBM system that enables the Datacomputer to store and access such a large volume of on-line data is described below. Our reliability experience with this mass storage system and a new utility program developed for it are them discussed.

5.1 General Description

The CCA Datacomputer has been equipped with the first public installation of the Ampex Tera-Bit Memory (TBM) System. This device uses video tape technology to achieve a maximum on-line capacity of around 3 trillion bits. Maximum seek time to any bit is 45 seconds. Maximum data transfer rate is 5.3 million bits per second.

The TBM at CCA is equipped with two dual tape transport modules so at most four tapes, or about 176 billion bits (equivalent to 220 IBM type 3330 disk packs) can be available on-line. All equipment between the transports and the Datacomputer central processor is non-redundant in

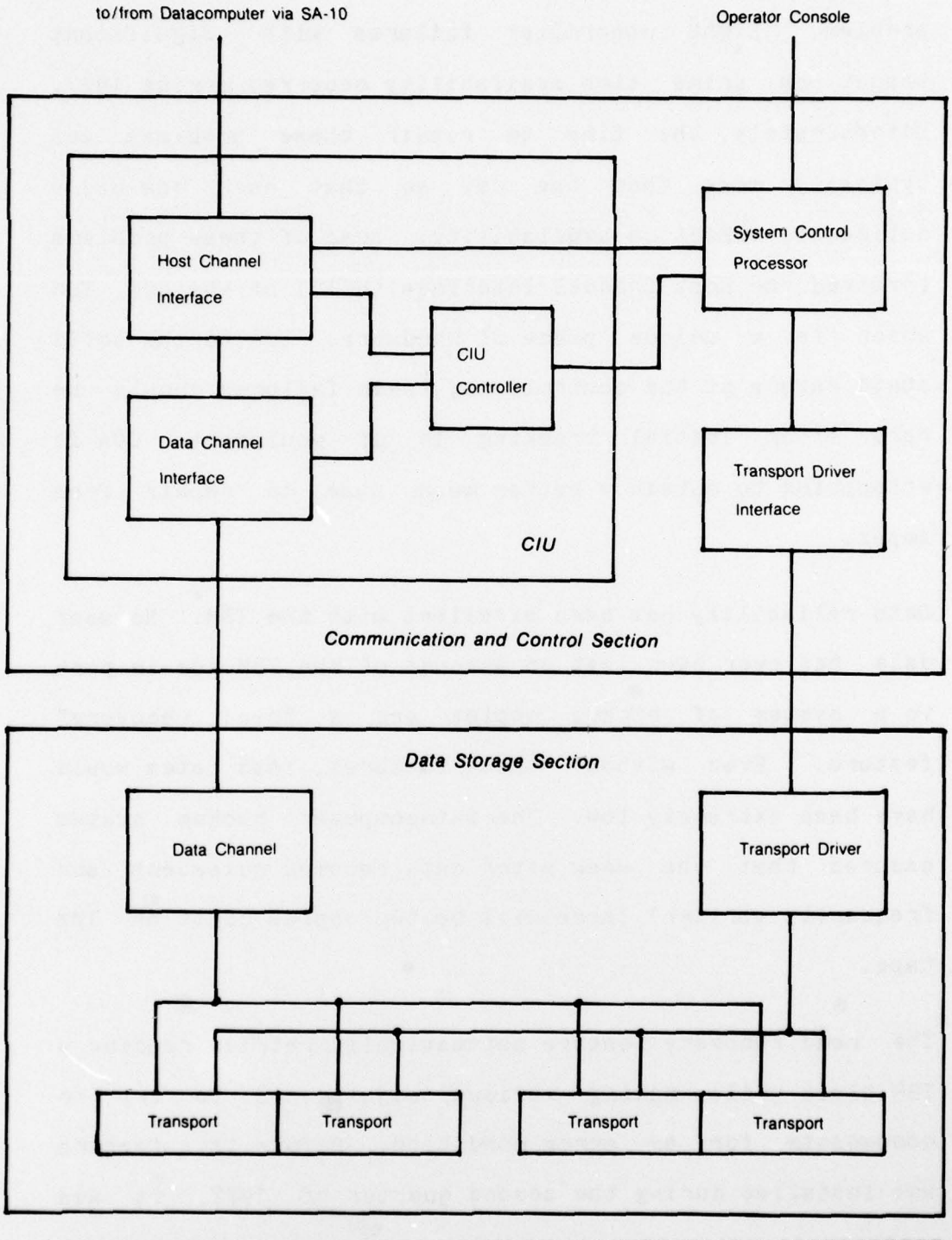
the present CCA configuration as shown in Figure 5.1. There is one transport driver (necessary for a tape to be in motion), one data channel (necessary to encode and decode digital information to and from the broadband signal on tape), one system control processor to coordinate the TBM, and one channel interface unit that connects the TBM system to the Datacomputer's PDP-10 system. Data is transferred directly between TBM tape and core memory.

5.2 Reliability

There are three different facets to TBM reliability. They are mechanical reliability, controller reliability, and data reliability.

Problems with mechanical parts of the TBM, such as magnetic head wear on the transports, have proven to be predictable and manageable through normal maintenance procedures. Mechanical components of the system are redundant in that there are multiple transports and backup air compressors and vacuum pumps on line. Spares are kept on site and no mechanical failure has put the system out of operation for a significant length of time.

CCA TBM Hardware Structure Figure 5.1



Controller reliability has been the worst TBM availability problem. Eight controller failures with significant impact on prime time availability occurred during 1977. Unfortunately, the time to repair these problems was typically more than one day so that each one had a noticeable impact on availability. Some of these problems involved the Host Channel Interface (HCIF) of the CCA TBM which is a unique piece of hardware. Due to the solid state nature of the controllers, their failures should be rare after initial breaking in of equipment. CCA is attempting to obtain a better mean time to repair from Ampex.

Data reliability has been excellent with the TBM. No user data has ever been lost on account of the TBM due in part to a system of backup copies and a "read recovery" feature. Even without these features, loss rates would have been extremely low. The Datacomputer backup system assures that one week after data becomes quiescent (and frequently earlier) there will be two copies of it on TBM tape.

The read recovery feature automatically retries reading a TBM block while making various adjustments to try to compensate for an error condition. Before this feature was installed during the second quarter of 1977, it was

necessary to try these adjustments manually when a block not normally readable was encountered. Although a rare occurrence (one the order of once every three weeks), this sort of manual intervention did interfere with normal operation. Furthermore, since the Datacomputer is frequently unattended, access to data to could be delayed for some time. Now, after normal reads fail, the Datacomputer can invoke the TBM read recover operation which takes a maximum of two minutes.

5.3 TBMUTL

In the past, a profusion of small and medium size TENEX programs were written for testing particular aspects of TBM operation, absolute dumping of information in readable form from TBM tape, verifying and testing correct overall TBM operation, demounting TBM tapes, reading TBM drive usage statistics, and similar utility operations. All these programs were written in machine language to meet particular needs.

This year, a unified TBM utility program, called TBMUTL, was developed to replace all of the earlier programs. For ease in development and future maintenance, TBMUTL was written in BCPL, a high level compiler language maintained

for TENEX by Bolt Beranek and Newman (BBN). A user manual was issued with the original version of TBMUTL.

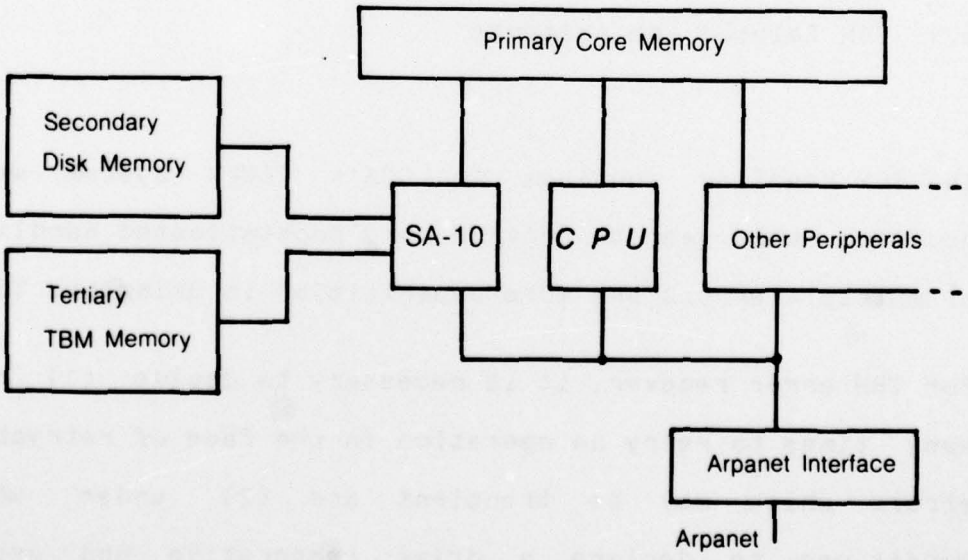
During the third quarter, TBMUTL was expanded in the following ways, and a complete revised user manual was issued:

1. a quick test sequence was added for checking out TBM drives after cleaning,
2. a printing suppression feature was introduced to make TBMUTL more usable on slow terminals,
3. commands were added for doing Read Recover and Auto Align operations,
4. TBMUTL was modified so that multiple drive tests could be done more conveniently and
5. a convenient top level command was added to read in and clear the operations counts for all TBM drives.

6. CCA TENEX

The Datacomputer runs as a user job under the CCA TENEX operating system. Many modifications had to be made in TENEX to accommodate the special requirements of the Datacomputer and the special hardware in use on the CCA Datacomputer. Figure 6.1 shows the hardware structure of the system.

CCA TENEX Hardware Structure Figure 6.1



Among the general modifications that had been made were

1. improved efficiency in the network interface code for the high rate of maximum size network messages send and received by the Datacomputer;

2. device code for using the Ampex TBM Mass Storage system and CalComp disks; and

3. additional statistics gathering code.

Enhancements made to CCA TENEX in 1977 are described below.

6.1 TBM Related Enhancements

The TBM handling routines in CCA's TENEX system were modified this year to provide more sophisticated handling of multiple errors and more capabilities in using the TBM.

For TBM error recover, it is necessary to decide (1) how many times to retry an operation in the face of retryable errors which may be transient and (2) under what conditions to declare a drive inoperative and avoid

referencing it. Based on our error experience, the number of attempts for retryable errors was reduced during the first quarter of 1977 to avoid excessively degrading system throughput in the rare intervals where irrecoverable errors are being encountered. The criterion for declaring a TBM drive down pending operator action, to avoid making things worse by pounding on a defective drive, was made considerably more complex. Errors were divided into two categories, fatal and suspicious. Fatal errors, such as those indicating that operator intervention is required, cause the drive to be declared down immediately. Suspicious errors are counted and the drive declared down after sixteen. Warning messages for a TBM drive are also counted and a drive declared down after a large number of them. (A retryable error that succeeds when retried does not increase any of these error counts.) When a tape is mounted or aligned, the error and warning counts are cleared.

Later in 1977, the burden of retrying on data failures for ordinary reads was transfered to the internal TBM software. At that time TENEX was modified not to retry ordinary reads and to treat their failure as it had previously treated a failure that exhausted the allowed number of retries.

CCA TENEX was also augmented with two system calls, one to issue a read recover command to the TBM, the other to do a "raw write". Read recover, which Ampex implemented this year in the TBM internal software, will automatically perform all the normal manual steps taken to recover a block of data for which difficulties in reading are being encountered. It is usually invoked after the normal read fails with the usual retry strategy. The raw write system call will simply write a block of tape. To be assured that a write is successful and achieve low error rates, it is always necessary to verify a block after it is written. All previous TBM write operations in TENEX did a write followed by a verify as a unitary operation. This provides the maximum reliability but it is significantly more efficient to write several adjacent blocks and do a continuous multi-block read verify. Raw write makes it possible to obtain this extra efficiency.

6.2 General Enhancements

Growing use of the Datacomputer and a growing Datacomputer directory increased the swapping and file load on the CCA TENEX disk system. To compensate for this the CCA TENEX file system was expanded to include another disk drive and reorganized to use part of each disk drive for swapping.

The disk drive which was added to those available for file storage had previously been reserved for backup in case of failure in another drive and for Datacomputer development. The addition of a disk drive represented an increase in space available for file storage by 25%.

Before reorganization of the disk allocation between file and swapping areas, disk transfers due to virtual memory swapping had all been concentrated on one drive. By using an area on each drive instead, the load is more evenly distributed. During heavily loaded periods, most transfers are for swapping and the disk heads tend to dwell in the limited swapping area on each drive reducing head motion time and improving response.

A. General Description of the Datacomputer

This is a brief general description of the structure of the Datacomputer which is intended to provide context for other parts of this report. Persons already familiar with the Datacomputer may skip over it. Persons desiring a more detailed description than presented here are referred to the final Semi-Annual Technical Report for the Datacomputer Project, contract number MDA903-74-C-0225, covering July through December 1976, and to the two papers, entitled Use of the Datacomputer in the Vela Seismological Network and Tertiary Memory Access and Performance in the Datacomputer, which are available from Computer Corporation of America as Technical Reports CCA-77-08 and CCA-77-11 respectively.

The intended Datacomputer user is a program running on a Arpanet host remote from the Datacomputer. This program calls the Datacomputer over the network and establishes a pair of standard uni-directional 8 bit byte network connections. The user program then proceeds to send Datalanguage over one connection while the Datacomputer

replies on the other. Datalanguage is a uniform high level language which gives the user a hierarchically structured view of his data.

Actually, the Datacomputer sends a "reply" to prompt the user program whenever the Datacomputer is ready to accept another line of Datalanguage. Similarly, the Datacomputer keeps the user program abreast of the progress of its requests with various synchronization, error, and success messages and comments. All replies have a fixed format which is designed for easy parsing by the user program. The reply begins with a number quickly identifying certain important messages. This is followed by the date, time, and an internal Datacomputer identification of the message source. The remainder of a reply provides a human-readable version of the message to be conveyed. In the case of serious errors, to assure resynchronization with the user program, the Datacomputer enters a mode where it rejects all messages from the user, giving an appropriate reply each time, until the user program sends a special message to clear this condition.

Data as well as commands and replies can be transmitted between the user program and the Datacomputer over these connections but certain characters are prohibited and the connections are fixed at an 8 bit bytesize. More general

data transfers can be done by using a separate data connection.

The requests, replies, and data for a particular user program are handled by the half of the Datacomputer known as RH or the request handler. RH handles the parsing of the user commands, which are in datalanguage, and synthesis of replies. For more complex commands, RH takes the user's requests and the data descriptions stored in the Datacomputer and compiles them, in several stages, into code to execute the request.

Data descriptions in the Datacomputer are associated with each file of data. Data descriptions are also provided for data streams to be received from or transmitted into the Arpanet. These streams are known as ports. Descriptions are set by the user when a file or port is created. Datacomputer data descriptions provide for hierarchical arrangements of structures of diverse data elements and repetitive lists. Many data types and data formatting alternatives are provided. This is important in ensuring that the Datacomputer can communicate efficiently with its diverse class of remote user computers.

RH accomplishes most of its activities by calling on routines in the other half of the Datacomputer, known as

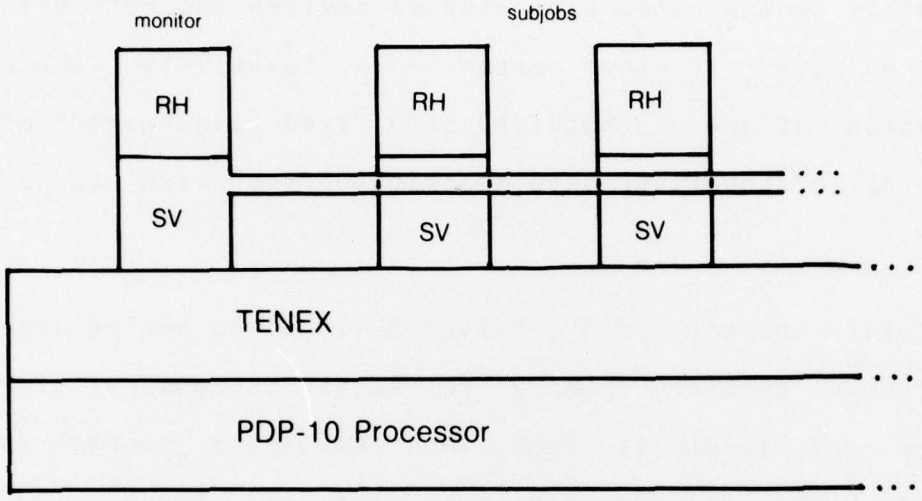
SV or services. SV is a pseudo-operating system in which RH runs. It is SV that actually has custody of the Datacomputer's multilevel hierarchical directory tree and enforces the extensive protection mechanisms of the system. SV is the part of the Datacomputer that ultimately communicates with storage devices and retrieves or stores bits. It views data as a relatively simple collection of areas subdivided into fixed size "pages" of data. All of the finer data description mechanism is in RH.

A special subpart of SV, called SDAX, moves active data from slower tertiary memory to faster secondary disk storage and moves it back when secondary storage is crowded. The CCA Datacomputer uses an Ampex TBM, described in Section 5 above, for tertiary storage. SDAX keeps track of where various copies of each file data page are. It also ensures data consistency by preventing updates of a file that are not successfully completed from affecting the original file. SDAX allows multiple readers and a single updater of a file among these Datacomputer subjobs. Each reader sees a consistent version of the file including only updates that were complete when they opened the file.

At any one time, there are multiple copies of the RH-SV pair serving users over the network as shown in Figure

---

Figure A.1



A.1. This actually means multiple copies of their variable area as they are implemented in reentrant code. The RH-SV pairs are called subjobs as they are all part of one job, under a monitor process, running in our modified TENEX operating system.