UTILIZING IBM PLUG-COMPATIBLE DISKS ON THE CDC CHANNEL

John W. Nall
Computer Center
Florida State University
Tallahassee, Florida

In August of 1977, the Florida State University was experiencing a severe problem with a lack of sufficient mass storage, with every sign that the problem would continue to get worse. In considering solutions, it was observed that IBM plug-compatible disks were being sold for about half the price of comparable CDC units. Unfortunately, the CDC channel would not accomodate these disks.

At about the same time, we found that a new company, Network Systems Corporation, of Minneapolis, Minnesota, was offering a new product which at that time they called "channel adapters" (later to be known as the "hyper-channel"). It was claimed that these adapters could be utilized to drive IBM plug-compatible disks from our CDC equipment.

After exploring the issue further, it was found that while the hardware might do this, there was no software available for the CDC computer to drive the adapters. After negotiations with NSC, we proceeded to acquire an A110 adapter (for hooking to the CDC channel) and an A510 adapter (for simulating an IBM channel), with the understanding that we would attempt to develop software to drive disks. We also acquired a Memorex 3672/3673/3675 disk subsystem, equivalent to a pair of IBM 3330 II drives.

At that time, the FSU computer center had a CDC 6400 and a CDC CYBER 73 computer. The 6400 was later swapped for a CYBER 74, and CYBER 73 later swapped for a CDC CYBER 170/730. We were operating under the KRONOS 2.1 Operating System. Control data hardware consists of a central CPU, with multiple small computers called "PP's" interfacing to I/O equipment over the CDC data channels. The CPU and PP's communicate only through central memory. Under the KRONOS Operating system, as well as under its successor, NOS, each PP does its own disk I/O (as opposed to the "stack processor" idea, where one PP does all disk I/O and is passed requests). In order to allow each PP to the facility of doing its own disk I/O, a small area is reserved within the PP called the "driver area. This area, along with a small area at the top of PP memory known as the "error processor area," comprises approximately 380 words decimal, which is all that is available for a mass storage driver.

We felt at the beginning of the project that the driver should be in every way transparent to the Operating System, such that any routine which called a mass storage driver should be able to call the driver for the IBM disks with no knowledge that they were not CDC disks. The rigidity imposed by this self-imposed restraint, together with the necessity of carving up the necessary code into multiple overlays in order to fit within the necessary space, resulted in a failure to meet our timing goal. The resulting driver was too slow. In particular, on a disk write, it was only making one disk revolution per write. That portion of the project consumed approximately 12 months, after which time we decided to modify our design.

The second design called for a radical departure from the CDC concept of disk storage, by utilizing huge sectors. This meant that it would no longer be possible to have a "transparent" driver. Abandoning this, however, gave us a great deal more freedom of design, and we decided that the disk system should be capable of being shared by several computers by means of a network of channel adapters all accessing a common disk subsystem.

There have already been a couple of papers written on this project, and there is a detailed set of documentation available. If anyone wishes to go into it at more depth than that provided by this talk, please contact me and I will be glad to provide whatever further information may be desired.

1.0.0 INTRODUCTION.

The Florida State University Network (FSUNET) is a local high-speed data communications network, utilizing Network Systems Corporation (NSC) adapters to inter-connect the components of the network.

The network is an open-ended network, in that any number of nodes may be added to the multi-drop communications net, by merely adding additional NSC adapters.

The only components of the network which are initially present are the CYBER 73, the NSC A110 adapter, the NSC A510 adapter, and the secondary mass storage capability of the Memorex disk system. See Figure 1.8.

Within this paper, the term 'NETEXEC' refers to a particular implementation scheme of the software module to be used in a CYBER computer, operating with the Kronos 2.1 Operating System, to implement a host computer interface to the network. This paper is meant to describe the general guidelines which must be followed by an analyst to write such a software module for any host computer to interface to the network, as well as to describe NETEXEC in detail.

1.1.0 GENERAL PHILOSOPHY

The Memorex disk under FSUNET is a secondary mass storage medium, utilized in much the same manner as indirect access permanent files under KRONOS 2.1. No user will be able to directly read/write the disk, but will only be able to ask that a file be copied to or from FSUNET.

It should be clearly realized in reading this paper that 'NETEXEC' refers to a subsystem on the CYBER 73 computer, operating under the Kronos 2.1 operating system, while the 'network storage system' is a part of the network, and no more dependent upon the CYBER than upon any other host computer which may be in the network. Thus, while of necessity some terms are used in referring to the network storage system which have a certain meaning within the CYBER terminology, these terms must be given their general meaning. Such terms are 'file', 'user index', 'character', 'control point' and others. In such instances, the user must read the term in its general sense, realizing that host systems other than the CYBER will have some analagous term.

1.2.0 DISK ALLOCATION.

The disk is allocated in terms of data sectors, rather than in terms of tracks as with CDC disks. Each data sector consists of 31200 bits of data information. (An exception to this is the label cylinder, discussed in more detail later, which will have 31296 bits of data information per data sector). It should be noted that this can be divided into either 5200 6-bit units, 3900 8-bit units, 1950 16-bit units, 1300 24-bit units or 975 32-bit units. This data information is further subdivided into 480 bits of header information, which may be looked at in terms of 80 6-bit units, 60 8-bit units, 30 16-bit units, 20 24-bit units, or 15 32-bit units, and 30200 bits of user-data.

Network storage is global, in the sense that it may be accessed by any host computer on the network. Host computers may be of varying word-sizes, and varying byte-sizes. The above discussion shows that many variations are allowed for in the network storage scheme. However, as we are presently oriented to the CYBER family, operating under Kronos 2.1, the remainder of this discussion will largely concern itself with discussing network storage in terms of 6-bit characters.

Each platter of a particular cylinder consists of three of these sectors, thus allowing 15600 characters per platter on a cylinder. (There are also three 'filler' sectors, one at the end of each data sector, in order to equally space the data sectors around the platter. This provides the maximum housekeeping time in between data sectors).

Each cylinder utilizes all 19 surfaces, thus allowing 19 x 15600, or 296,400 characters per cylinder. 814 cylinders will be utilized, thus allowing 814 x 296,400, or 241,269,600 characters per disk unit. Of these, 237,557,760 will be actual data -- the remainder will be header information (the 80 characters at the beginning of each sector). One cylinder will be used for the pack label, the Sector Reservation Table (SRT), and the File Access Table (FAT).

(It should be noted that the 844 dual-density drive contains a total of 238,259,200 characters. Thus the Memorex disk will be .997 capacity of the dual-density 844).

1.2.1 SECTOR LINKAGE.

A 'file' on network storage refers to all that information which may be accessed by a single user who has furnished one name by which the 'file' is known, i.e., a collection of one or more data sectors , referred to by a 'file name'.

The one or more data sectors which make up a file must all be contained on one disk pack, i.e., there is no pack overflow provision. When assigning disk space for a user, NETEXEC will assign consecutive data sector numbers contained on one cylinder, where space permits. This unit of one or more consecutive data sectors is referred to as a 'consec'. There may be more than one consec on one cylinder, i.e., a user may be assigned the first 7 data sectors on a cylinder, and the last 5 data sectors on the same cylinder. If there is not sufficient space on one cylinder for the file, consecs will be assigned on another cylinder, which need not be a consecutive cylinder, although it may be.

The header of each data sector of each consec points to the next sector, which may or may not be on the same cylinder. The header of the last sector in the file will so indicate. See Figure 1.4, word 2.

It should be noted that nothing in the above definition prohibits a consec from consisting of only one sector. The sector allocation algorithm is set up to assign some minimum number of sectors as a consec. It may turn out that analysis shows that a consec may consist of only one sector without a problem.

1.3.0 SECTOR ALLOCATION

Since only copies are performed, the Chippewa Operating System sector allocation scheme is used, where each bit in a word represents a sector, and is "on" to indicate reservation. This scheme did not work well as a general allocation scheme under Chippewa, because of the backspacing problem, but this will not be a problem under FSUNET since no backspacing will take place. (In Chippewa, sectors were linked on the disk, but not linked in any central memory table. If a user needed to backspace into the previous sector, it was necessary to read the whole file from the beginning to determine which sector was the previous sector, since sectors were linked in a forward fashion).

1.4.0 MINIMUM HARDWARE.

In order for the network scheme to become operational, a certain minimum requirement of hardware must be met. Tentatively, this hardware consists of one A510 adapter, one disk pack and associated controller, and one host adapter (an A110 for the purposes of NETEXEC).

1.5.0 CHANNEL PROGRAMS

Because of the relatively large space between disk sectors, corresponding to about 5.1 milliseconds, it will be sufficient to have only one read or write per channel program, since time will exist during which another STARTIO can be issued. NETEXEC will notify the PP when the final sector is being written, and the PP will be aware when the final sector is read, thus the PP knows not to issue another STARTIO in that event.

1.5.1 CHANNEL PROGRAM FORMAT.

The channel command words (i.e., A510 programs) are contained within STL (overlay 3DM). The A510 is loaded by STL at deadstart time if the EST contains a DN device. The A510 may also be loaded by calling the PP program *NDN* from the console (under DIS). *NDN* merely calls the 3DM overlay and executes it.

There are four (4) channel command programs. CCW1 is the program which performs data reads. CCW2 performs data writes. CCW3 performs reads from the label cylinder. CCW4 performs writes on the label cylinder. The following is the format of each program.

SEEK (command code 7 hex)
SEARCH ID EQUAL (command code 31 hex)
TRANSFER IN CHANNEL (command code 8 hex)
READ DATA (command code 6 - 3900 bytes of data)

SEEK
NOP WITH PCI (command code 3)
SEARCH ID EQUAL
TRANSFER IN CHANNEL
WRITE DATA (command code 5 - 3900 bytes of data)

SEEK
SEARCH ID EQUAL
TRANSFER IN CHANNEL
READ DATA - 3912 bytes of data

SEEK
NOP WITH PCI
SEARCH ID EQUAL
TRANSFER IN CHANNEL
WRITE DATA - 3912 bytes of data

1.6.0 DISK LAYOUT.

Each platter has, in addition to the Home Address record ("HA") and Record Zero ("R0"), six data records per cylinder. These are designated as R1, F1, R2, F2, R3 and F3. (See Appendix 1.8). User data will only be written on R1, R2 and R3. F1, F2 and F3 are for the purpose of providing a "filler space" between the records in order to locate the user records at equal distances around the circumference of the platter. This provides the maximum time for activity by the driver between data transfers to or from the disk.

1.8.0 NETEXEC INTERACTION WITH FAT.

The File Access Table (FAT) resides on cylinder 400 of each device, and contains a three-word entry for each file on the device.

NETEXEC must both make entries in, and delete entries from the FAT. The PP routine 1SN will never do anything with the FAT other than to read or write (a sector at a time) on command from NETEXEC, or to search the FAT for a file.

NETEXEC utilizes the 'Search FAT' task (SFT) to have 1SN search the FAT for a file. The search by 1SN is end around, i.e., 54, 55, 56, 2, 3, . . . , 53.

2.0.0 1SN - PP ASSISTANT FOR NETEXEC.

1SN is the PP assistant which provides I/O capability for NETEXEC. 1SN's primary function is to perform 'tasks' which are given to it by NETEXEC. (See the common deck COMSDND for an explanation of these various tasks).

1SN is limited in space because of the extraordinarily large sectors which are used on the mass storage disk. These sectors are 5200 characters in length; a sector will occupy 2600 (5050 octal) PP words. This extremely large buffer makes it necessary to split 1SN into many overlays. These overlays must be kept within CM at all times, as they are loaded into the mass storage area (origin at MSD, or 600 octal).

2.1.0 1SN MEMORY USAGE

The sector buffer begins at 2700 octal in PP memory. Space below that is used by overlays (which origin at 600 octal and cannot go past 1073 octal) and the permanent 1SN resident which begins at PPFW (1100 octal). The permanent 1SN resident consists of the code which must identify a task and load the necessary overlay, and subroutines which are utilized by more than one overlay.

3.0.0 1SN PP RESPONSES.

NETEXEC must be preprogrammed to respond exactly to each PP-response from 1SN, according to what task is being performed. That is, the response which NETEXEC should make to a PP-response, is a function not only of the PP-response itself, but also of the task that is being performed. These responses are explained at more length in the expanded paper.

4.0.0 WRITING AND READING FILES.

4.1.0 WRITING FILES:

NETEXEC must set up a standard FET, which is pointed to by the job buffer. NETEXEC must also issue the necessary preliminary tasks to 1SN, prior to attempting to write a file to network storage.

Once the buffer has been filled, a WDF task will be issued to 1SN by NETEXEC. One WDF task will cause 1SN to write the entire file without a further task being issued. However, there is constant interaction between 1SN and NETEXEC during the file copy, in order to keep the data buffer replinished. Whenever 1SN issues the PP response of UPW to NETEXEC, NETEXEC must then update the "out" pointer in the FET. NETEXEC may also have to call CIO to read more data into the buffer (this can be done while 1SN is writing.) NETEXEC must also update word two (2) of the Processor Control Area each time, modifying the logical sector number, the word count (512 on all except last sector), the physical sector (NETEXEC has the SRT to obtain that information) and the EOF/EOI field. Note that this is the only information 1SN receives as to whether or not this is the last sector.

When 1SN finally issues the PP response of WFB to NETEXEC, the wait bit will be set to allow NETEXEC to do any needed cleaning up before assigning the next task. read). The data at that point is irrecoverable.

5.0.0 NETEXEC INTERACTION WITH SRT.

Track assignment, by way of setting or clearing bits within the SRT, is the exclusive responsibility of NETEXEC. The PP program 1SN only serves as an I/O device to read or write the SRT; it will never set or clear a bit within the SRT.

Word 400 (decimal) of the SRT is always reserved, as this is the label cylinder, which contains the label, the SRT and the FAT for the device. The remainder of the SRT is available, unless flaw bits are set. Flaw bits will always be set at the individual logical sector level, rather than at a cylinder level, unless the whole cylinder is unavailable.

6.0.0 TABLES.

The remainder of the paper consists of the different tables which are utilized by the system. Here again, the extended paper deals at more length with these than is possible here, and it should be obtained by anyone who wishes to pursue the subject.

FILE TRANSFER BUFFER

59 48 18 17 0

LOGICAL FILE NAME (FOR NETGET, NETPUT, NETREP)		USER INDEX OF USER AT NETHELP
CODE WHICH NETHELP PASSES		ADDRESS OF STATUS WORD WITHIN NETHELP
JOBNAME OF NETHELP JOB		

Fig I.1

FILE ACCESS TABLE (FAT)

ENTRY

59 17 0

LOGICAL FILE NAME				USER INDEX FILE WAS SAVED UNDER	
FILE SIZE IN SECTORS	FIRST SECTOR NUMBER	~~origin~~ device number	adapter number		
number of accesses	date file was created	date file last modified		date file last accessed	

Fig I.2

NOTE: Although the FAT consists of three 60-bit words per entry within the CYBER memory, it is written to network storage as three 64-bit words, utilizing the 60:64 assembly/disassembly mode of the A110 adapter.

PROCESSOR CONTROL AREA

59 17 0

0	LOGICAL FILE NAME ON NETWORK STORAGE					USER INDEX ON DISK FILE
1	# of accesses	DATE LAST MODIFIED	DATE LAST ACCESSED			DATE FILE WAS CREATED
2	unused	HEAD CYLINDER SECTOR	WORD COUNT	EOI		HEAD CYLINDER SECTOR
3	*ACCESS INFORMATION					
4	*****RESERVED FOR FUTURE USE****					
5	*****RESERVED FOR FUTURE USE****					
6	*****RESERVED FOR FUTURE USE****					
7	*****RESERVED FOR FUTURE USE****					
10	**TASK POINTER WORD					
11	JOBNAME OF NETHELP JOB					USER INDEX OF NETHELP JOB
12	ADAPTER NUMBER	CYLINDER NUMBER	TRACK NUMBER (HEAD)	PHYSICAL SECTOR NO.		DEVICE NUMBER
13	LOGICAL FILE NAME ON NETWORK STORAGE (NAME THAT IS IN FAT)					ADDRESS OF FET IN NETEXEC CM
14	unused	ADDRESS OF LABEL BUFFER IN NETEXEC CM	ADDRESS OF FAT BUFFER IN NETEXEC CM			ADDRESS OF SRT BUFFER IN NETEXEC
15		ADDRESS OF STATUS WORD WITHIN NETHELP	ADDRESS OF SINGLE SECTOR BUFFER			ADDRESS OF BUFFER FOR MESSAGE (GMM)
16	*****RESERVED FOR FUTURE USE****					ADDRESS OF ERROR MSG FOR *ANH*
17	*****RESERVED FOR FUTURE USE****					

* ACCESS INFORMATION IS EXPLAINED IN MORE DETAIL IN PARAGRAPH OF THE TEXT.

** THE TASK POINTER WORD IS EXPLAINED IN MORE DETAIL IN

*****THE RESERVED WORDS ARE SET ASIDE FOR FUTURE EXPANSION OF THE SYSTEM. FOR EASE IN PROGRAMMING, THE ENTIRE 'PCA' NEEDS TO BE A POWER OF TWO WORDS IN LENGTH.

Fig 1.3

Sector Header

WORD 2 is changed each sector. The others remain the same.

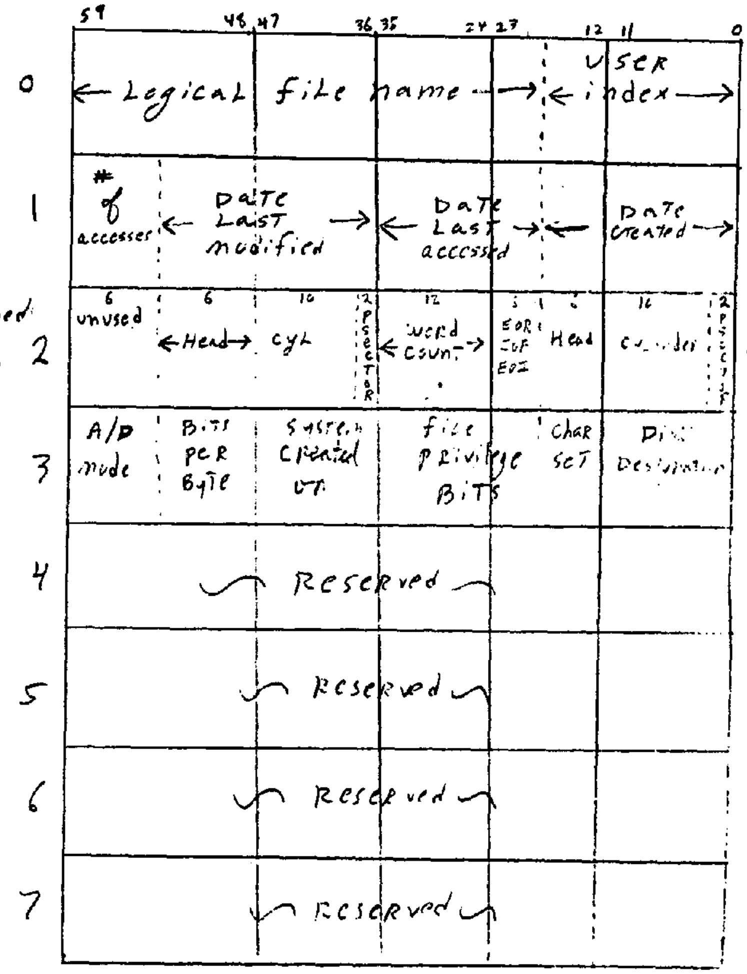

Fig 1.4

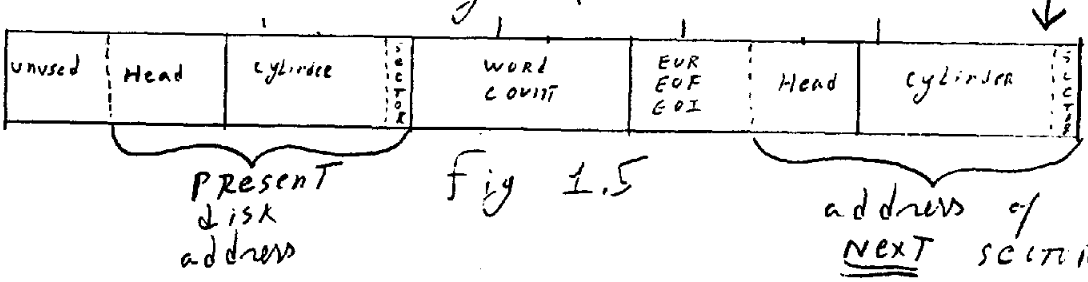

Fig 1.5

TASK POINTER WORD

59 58	46	41 ~~35~~ 24	17	0
PP-RESPONSE RETURNED BY 1SN TO NETEXEC	unused	ADDRESS OF CURRENT TASK WHICH 1SN WAS PERFORMING AT TIME WAIT BIT WAS SET	unused	ORIGINAL ADDRESS OF TASK WHICH 1SN WAS CALLED TO PERFORM, OR ADDRESS OF TASK GIVEN BY NETEXEC BEFORE THE WAIT BIT CLEARED.

FIGURE 1.6

TASK WORD

47 36	35	17
TASK CODE (USED BY NETEXEC FOR CODE TO USE WHEN SET WAIT BIT DUE TO A 'WAT' TASK.	unused	ADDRESS OF THE NEXT TASK TO BE PERFORMED (AFTER THIS ONE IS COMPLETE)

FIGURE 1.7

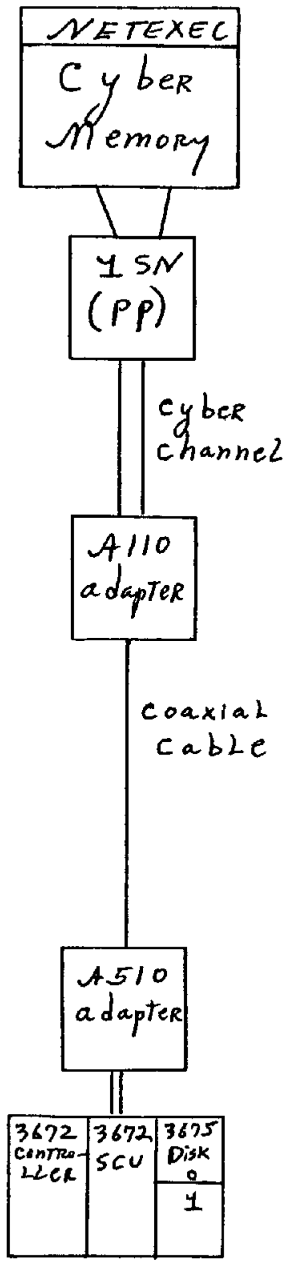
NETEXEC
Cyber Memory
1SN (PP)
cyber channel
A110 adapter
coaxial cable
A510 adapter
3672 Controller
3672 SCU
3675 Disk
0
1

Figure 1.8

SECTOR RESERVATION TABLE (SRT)

Word	59	56 ... 0
0		
2		
3		
4		
5		

THE SRT IS A TABLE, KEPT IN CENTRAL MEMORY OF NETEXEC AND ALSO KEPT (PERMANENTLY) ON THE DEVICE TO WHICH IT REFERS.

EACH SRT CONSISTS OF 816 WORDS. WORDS 0 THROUGH 814 CORRESPOND TO CYLINDERS OF THE DISK, AND WORD 815 IS A CHECKSUM OF THE OTHER 815 WORDS.

IN EACH WORD OF THE SRT, BITS 0 THROUGH 56 CORRESPOND TO THE 57 SECTORS OF EACH CYLINDER. A BIT IS 'ON' TO INDICATE THE SECTOR IS RESERVED.

THE REMAINING 3 BITS OF EACH WORD IN THE SRT (BITS 57 THROUGH 59 ARE UNUSED).

Word	59	56 ... 0
813		
814		
815		

Fig 1.9

SECTOR 0 AND SECTOR 1 OF CYLINDER 400 (decimal)

THE FIRST 80 CHARACTERS, OR FIRST 8 CM WORDS, ARE THE 'LABEL' OF THE DISK. THIS IS WRITTEN ANEW EACH TIME THE SRT IS WRITTEN TO THE DISK.
THE NEXT 154 WORDS ARE UNUSED AT PRESENT. THIS IS SO BECAUSE IT IS DESIRED THAT THE 816 WORDS OF THE SRT BE CONSECUTIVE.
THE REMAINING 327 WORDS OF SECTOR 0 COMPRISE THE FIRST 327 WORDS OF THE SRT.
THE REST OF THE SRT (489 WORDS) IS CONTAINED IN SECTOR 1.

Fig 1.10