AD-787 677

# DATACOMPUTER PROJECT

Computer Corporation of America

Prepared for:

Defense Supply Service
Advanced Research Projects Agency

30 June 1974

AD787677

Computer Corporation of America
575 Technology Square
Cambridge, Massachusetts 02139

# DATACOMPUTER PROJECT
# SEMI-ANNUAL TECHNICAL REPORT

March 13, 1974 to June 30, 1974

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by the U.S. Army Research Office, Defense Supply Service--Washington under Contract No. MDA903-74-C-0225. 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 Defense Advanced Research Projects Agency or the U.S. Government.

ia

## Abstract

The datacomputer system is being designed as a large-scale data storage utility to be accessed from remote computers on the Arpanet and, potentially, on other networks. The development is phased, with each successive release of the system offering increased capabilities to users. During the present reporting period, the second major release of the system became operational. This release, while still primitive in many respects, is beginning to provide experience with actual applications and user programs.

Table of Contents

## 1. Overview

### 1.1 Review of Basic Concepts

The goal of the project continues to be the development of a shared, large-scale data storage utility, to serve the needs of the Arpanet community.

The system under development will make it possible to store within the network such files as the ETAC Weather File or the NMRO Seismic Data File, which are measured in hundreds of billions of bits, and to make arbitrarily selected parts of these files available within seconds to sites requesting the information. The system is also intended to be used as a centralized facility for archiving data, for sharing data among the various network hosts, and for providing inexpensive on line storage to sites which need to supplement their local capability.

Logically, the system can be viewed as a closed box which is shared by multiple external processors, and which is accessed in a standard notation, "datalanguage" (see Fig. 1). The processors can request the system to store information, change information already stored in the system, and retrieve stored information. To cause the datacomputer to take action, the external processor sends a "request" expressed in datalanguage to the datacomputer, which then performs the desired data operations.

From the user's point of view the datacomputer is a remotely-located utility, accessed by telecommunications, It would be impractical to use such a utility if, whenever the user wanted to access or change any portion of his file, the entire file had to be transmitted to him. Accordingly, data management functions (information retrieval, file maintenance, backup,

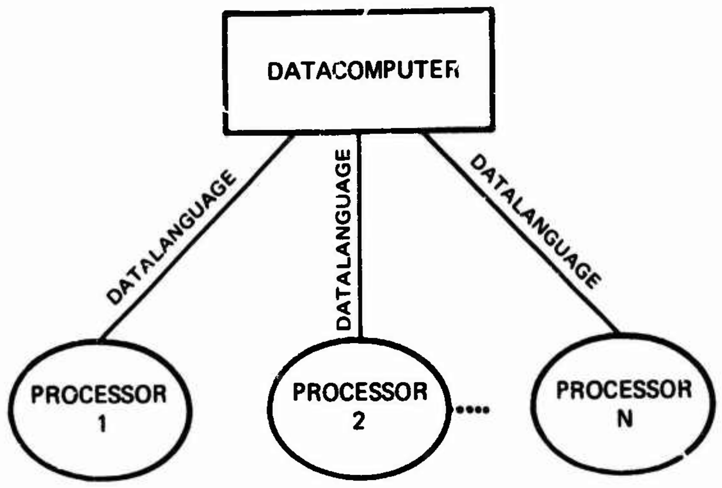

Figure 1. Logical View of Datacomputer

access security, creation of direct and inverse files, maintenance of file directories, etc.) are performed by the datacomputer system itself. The user sends a "request", which causes the proper functions to be executed at the datacomputer without requiring entire files to be shipped back and forth.

The hardware of the system is shown in overview in Fig. 2 and in greater detail in Fig. 3.

The program for the system processor handles the interactions with the network hosts and is designed to control up to three levels of storage: primary (core), secondary (disk), and tertiary mass storage. Currently, the CCA facility is operating with primary and secondary storage only, with the addition of tertiary storage planned for 1975. Installation of a tertiary storage module will leave datalanguage unchanged, and will therefore be imperceptible to users of the system (except insofar as it affects performance and the total storage capacity available for data).

In addition to using the dedicated equipment at CCA, it is planned that datacomputer service will also make use of hardware resources located at NASA/Ames, using CCA software. The two sites will provide mutual backup for one another, thereby guarding against accidental loss of data and providing for satisfactory uptime of the overall service.

## 1.2 Status of Project

During this reporting period, Version 0/10 of the datacomputer system was completed. This is the second major version of the system to offer datacomputer services on the Arpanet. Version 0/10 has replaced Version 0/9.7, which was an "intermediate" release. Version 0/10 handles non-ASCII and variable-length data. It has file-level access regulation. (See chapter 2 and Appendix for details.)

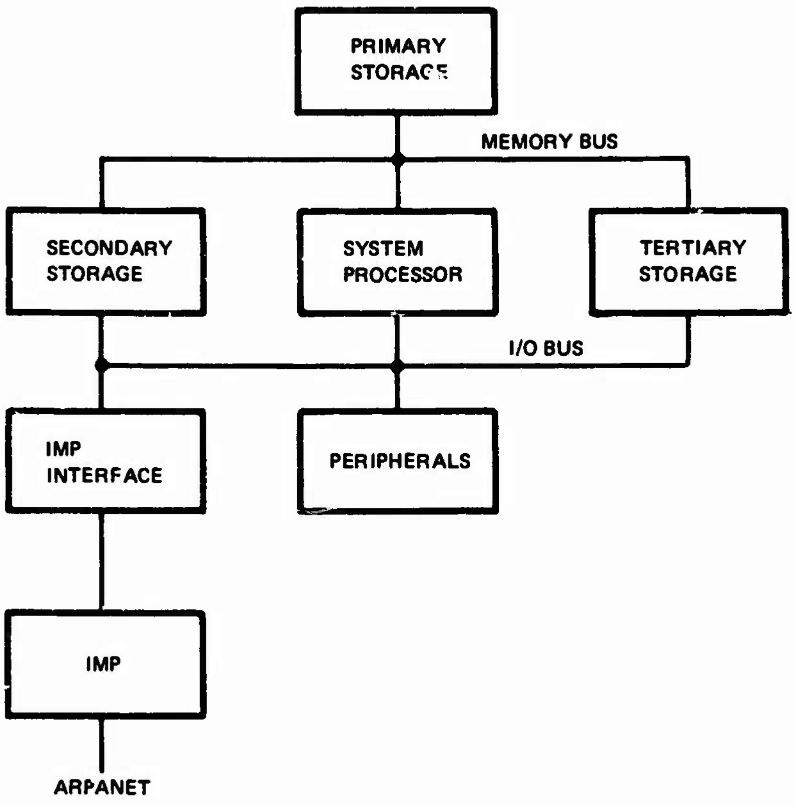

Figure 2. Hardware Overview of System

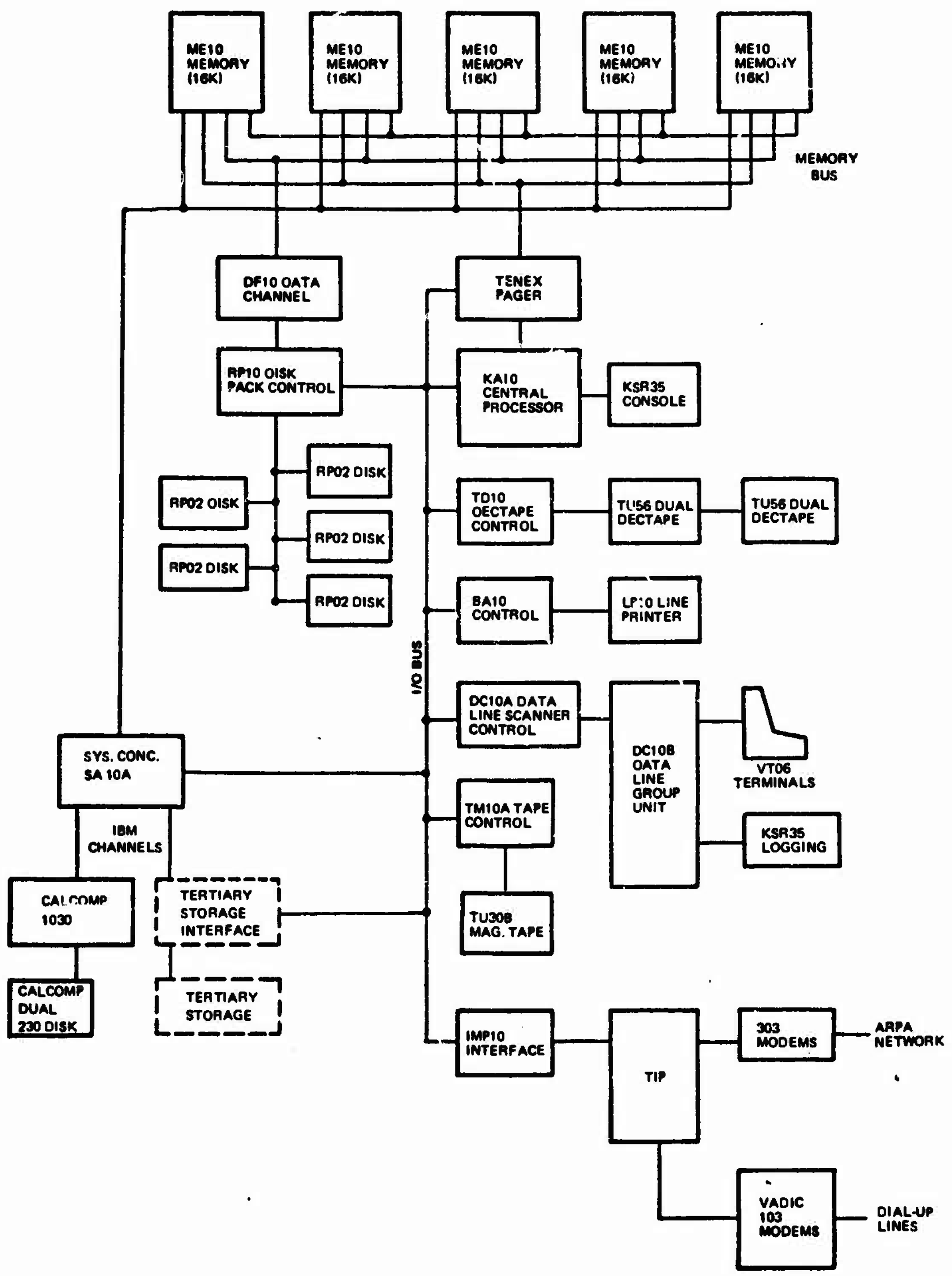

Figure 3. Hardware Block Diagram - CCA Installation
(Equipment in dashed outline is planned for 1975)

The project continues on an increasing scale to interact with actual and potential datacomputer users. New user programs have given us more operational experience with the system. Much attention is being paid to the seismic community, the weather community, and other users to determine their data-computer requirements and adjust the implementation priorities accordingly.

Currently only disk storage is available to the system. A Calcomp Dual 230 disk was installed during the fourth quarter of 1973. A second Calcomp Dual 230 disk will be added later in 1974. This will bring the total CCA storage capacity to about 4 billion bits. Plans call for the addition of large tertiary storage in 1975.

## 2. Software Implementation

During this reporting period, Version 0/10 replaced Version 0/9.7 as the system offering service of the Arpanet. The new features of 0/10 are summarized in this section. (See Appendix, "Datacomputer Version 0/10 User Manual" for details.) Specifications and implementation of Version 0/11 were begun.

### 2.1 Request Handler

Data Description. The datalanguage user must supply descriptions for all data, whether it is data being transmitted to or from the datacomputer (port description) or data being stored at the datacomputer (file description). The data may be tree-structured. The simple data types handled by Version 0/10 are 7-bit ASCII, 8-bit ASCII, and uninterpreted bytes or byte strings (with a user-specified byte size less than or equal to 36). Variable-length data may have either a one-byte preceding count, a one-byte delimiter, or, if it is in a port, a trailing "punctuation" character (i.e., end-of-record, end-of-block, or end-of-file marker).

The previous restriction that a file or port must be a list has been lifted. Also lifted is the size restriction that inner containers (i.e., containers inside of files or ports) must be less than 2560 characters.

The data description facilities in Version 0/11 will be the same as in Version 0/10.

Data Operations and Access Methods. In Version 0/10 the user may store files, retrieve files, replace files, and append to files. The user may also retrieve subsets of a file specified by boolean expressions on multiple variables. In retrieving or storing data, the datacomputer can also reformat it.

Unlike earlier versions, 0/10 allows members of inner level lists to be used in boolean expressions. (This is sometimes called a keyword feature; it allows an attribute -- or container, in datalanguage terms -- that occurs several times in one container with different values to be used in a retrieval specification.) Members of inner level lists may also be inverted. However, only EQ can be evaluated using the inversion; evaluation of NE still requires sequential search of the data.

Version 0/11 will introduce a rudimentary updating capability. Replacement of uninverted fixed-length containers will be possible. This includes fixed-length containers that are inside of variable-length containers.

In addition to specifying a set of containers by content, in 0/11 the user will be able to specify a set of containers by position in a list, called the index number of the list member. The set specification may be used either for retrieval or for updating.

In order to allow efficient retrieval of variable-length containers, the auxiliary structure, called a Container Address Table (CAT), will be implemented. The CAT provides a mapping from index number or internal record number to logical address. It can be used both for indexed and inverted retrievals.

Data Privacy. Version 0/10 has directory-level access regulation, that is, regulation at the file level and higher. The classes of users are defined by knowledge of passwords, by host, and/or by socket number. The privileges to be granted or denied are read, write, append, login, and control of privileges.

## 2.2 Services

Version 0/10 supports multiple volumes. This allows the datacomputer to use both of the CCA 3330-type spindles for storing datacomputer files. These disks are treated as "special disks", not as part of the normal Tenex page space.

A utility routine that dumps datacomputer files to magnetic tape was added. It can run as a background job without interfering with datacomputer services.

## 3. Network Services

### 3.1 User Programs

The datacomputer is accessed by user programs which run on other hosts on the Arpanet and send datalanguage requests to the datacomputer. In order to gain operational experience with the datacomputer and the problems associated with using it, and in order to facilitate usage of the datacomputer system, CCA has written a number of user programs.

During the previous reporting period, two such programs were written: SMART, which generates datalanguage for users at terminals, and FORPAC, which provides an interface between Fortran programs and the datacomputer. Based on our experience with these two programs, a set of standard subroutines (DCSUBR) needed for communication with the datacomputer were specified and implemented. There are routines to set up network connections, send datalanguage, send data, read data, and the like. Written in Macro-10, DCSUBR serves as a model for similar programs to be written for other machines.

One of the user programs to incorporate DCSUBR is RDC (Run Datacomputer). RDC provides convenient terminal access to the datacomputer from a Tenex host. Datalanguage is transmitted from either a teletype or from a local Tenex file, and datacomputer responses are displayed.

At the user's request, RDC will set up a secondary network connection as a data path to or from the datacomputer. This allows for transfer of non-ASCII data (not accepted over the datalanguage port) and it results in more efficient data transfers over the network. Unlike other user programs, RDC does not generate datalanguage; rather it gives a person a way to submit his own datalanguage. (The only exceptions are the datalanguage CONNECT and DISCONNECT statements.) RDC has been useful for debugging and for setting up new data bases.

A second program to utilize DCSUBR is DFILE. DFILE, which runs on any Tenex host, allows local users to archive their files on the datacomputer. The user, from his terminal, can associate attribute-value pairs with his file, and, later, retrieve the proper files based on boolean combinations of these pairs. DFILE may be used advantageously for files whose usage is not limited to a single host or for files which are public and meant to be distributed. The attribute-value pairs give the DFILE user a way of browsing through the DFILE database to find out what files are available.

## 3.2 User Statistics

The following chart indicates the number of times each network site has connected to the datacomputer in the present reporting period. During the period March through May, both Versions 0/9.7 and 0/10 were available over the Arpanet. The figures for this period indicate the changeover from one version to the next.

| | Total | CCA | MIT--DMS | Harvard | Other |
|---|---|---|---|---|---|
| January | 419 | 160 | 64 | 150 | 45 |
| February | 1323 | 143 | 1005 | 118 | 57 |
| March | | | | | |
| 0/9.7 | 1002 | 323 | 457 | 149 | 73 |
| 0/19 | 15 | 15 | | | |
| | 1017 | 338 | | | |
| April | | | | | |
| 0/9.7 | 956 | 284 | 381 | 174 | 117 |
| 0/10 | 391 | 350 | 20 | 0 | 21 |
| | 1347 | 634 | 401 | 174 | 138 |
| May | | | | | |
| 0/9.7 | 789 | 46 | 517 | 200 | 26 |
| 0/10 | 1142 | 712 | 186 | 191 | 53 |
| | 1931 | 758 | 703 | 391 | 79 |
| June | | | | | |
| 0/10 | 995 | 544 | 160 | 168 | 123 |

Number of CONNECTS to CCA
Datacomputer System - 1974

APPENDIX

Working Paper No. 9

"Datacomputer Version 0/10 User Manual"

June 1, 1974

# Datacomputer Version 0/10 User Manual

Datacomputer Project
Working Paper No. 9
June 1, 1974

Contract No. MDA 903-74-C-0225
ARPA Order 2687

A-14

**Computer Corporation of America**
575 Technology Square
Cambridge, Massachusetts 02130

# Datacomputer Version 0/10

# User Manual

Computer Corporation of America

1 June 1974

A-15

## Table of Contents

Appendices

# Chapter 1: Introduction to the Datacomputer

## Introduction

The datacomputer is a shared large-scale data utility system designed to serve the computers on the ARPA network. It may be thought of as a "black box" that performs data storage and retrieval functions in response to commands phrased in a standard notation, called datalanguage.

This document describes the currently-running version of the datacomputer software, and includes information about how a user program can access the system, transmit datalanguage, process the datacomputer's responses, and transmit and receive data over the network.

The datacomputer in its full implementation will provide an on-line storage capacity of one trillion bits and an extensive set of services to user programs. (1) The present version is a preliminary version, providing a limited amount of storage and a restricted set of user functions. Subsequent versions will progressively enlarge the range of services and the amount of storage available for users.

- - - - - - - - - - - - - - - - - - - -

(1) See Datacomputer Project Working Paper No. 8, Further Datalanguage Design Concepts, December 1973.

# Chapter 2: Containers

## Containers

The container is a basic concept in datalanguage. A container is an imaginary box which, like a FORTRAN variable, may contain data; a container may also enclose other containers. For example, some information about people could be represented as:

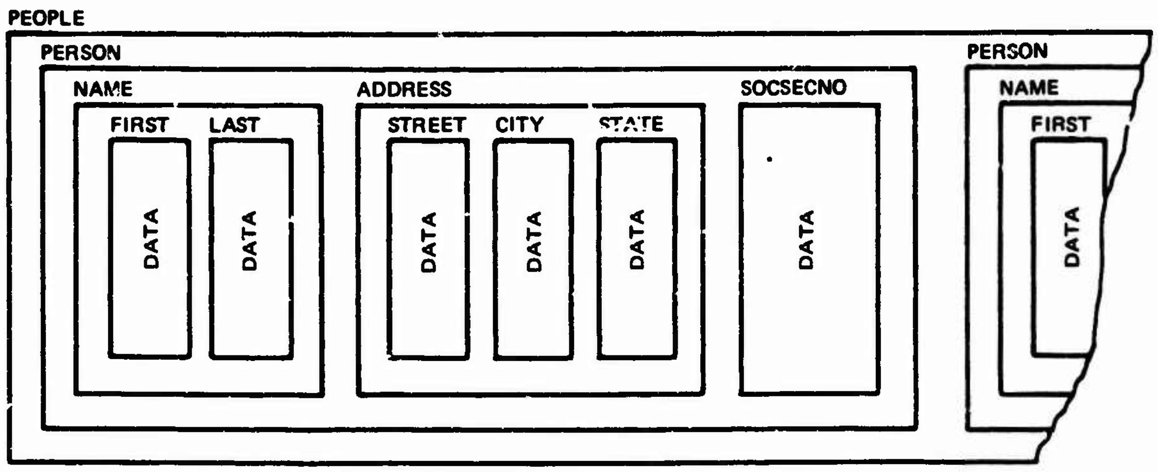

Figure 2-1. A container structure

Here PEOPLE, PERSON, NAME, and ADDRESS are containers enclosing other containers; FIRST, LAST, STREET, CITY, STATE, and SOCSECNO are containers that enclose only data.

The description of a container has several parts. It includes the container's ident, type, and size, and perhaps some additional attributes. The container's ident, or simple name, is a string of 100 or fewer letters, digits or the special character %, by which datalanguage requests refer to the container. The first character of an ident must be a letter or the character %. Certain reserved words may not be used as container idents; these are listed in Appendix B of this document.

Some sample idents are:

```
AVERYLONGIDENTABCDEFGHIJKLMNOPQRSTUVWXYZ
PEOPLE
WEATHER0STATIONS
%CCA
```

Containers are of four _types_, depending on their contents.

A container that is a _LIST_ contains some number of other containers. The LIST-members may be containers of any data type, but they must all have the same description. PEOPLE (above) is an example of a LIST.

A container that is a _STRUCT_, or _STRUCTURE_, contains some number of other containers, which need not have identical descriptions.(1) The descriptions of all the containers that are enclosed by the STRUCT form part of the description of the STRUCT itself; and on every occurrence of the STRUCT every one of its sub-containers must appear in the same order. ADDRESS is an example of a STRUCT.

A container of type _BYTE_ contains one byte of data. A container that is a _STR_ or _STRING_ contains a string of bytes.(2) FIRST is an example of a STR. The user can specify the byte size of BYTEs and STRs and can indicate an interpretation of 7- or 8-bit ASCII or uninterpreted (See below).

A LIST or STR has a _size_ associated with it. The size may be fixed or variable. The size of a STR is the number of characters in it, while the size of a LIST is the number of elements in the LIST.

## _Outermost Containers_

A container that is not contained by any other container is called an _outermost container_; outermost containers are different in several respects from other containers.

An outermost container in datalanguage has a _function_, which is either FILE, PORT, or TEMPORARY PORT (which may be abbreviated TEMP PORT). A FILE contains data kept in the datacomputer. When a FILE is created (see below), datacomputer space is allocated for it. A PORT describes data that is transmitted to or from the datacomputer. A TEMP PORT is a PORT whose description is not permanently

- - - - - - - - - - - - - - - - - - - -

(1) STRUCT and STRUCTURE are synonyms in datalanguage. Hereafter, STRUCT will normally be used.

(2) STR and STRING are synonyms in datalanguage. Hereafter, STR will normally be used.

stored, unlike the descriptions of other containers. TEMP PORTs vanish at the end of the session in which they were created.

## The Directory

The ident of an outermost container, whether it is a FILE or a PORT, is unlike other idents, in that it is entered in the datacomputer's directory. The directory is conceptually a tree; the entries in it are called nodes. A node may have one or more subordinate nodes, unless it represents a container, in which case it cannot. A portion of a hypothetical directory is diagrammed below; it may be read as indicating that the nodes F and G are subordinate to DATA, which in turn is subordinate to CCA. Only the bottom-most nodes in this tree, F and G, may represent containers, and they represent outermost containers.

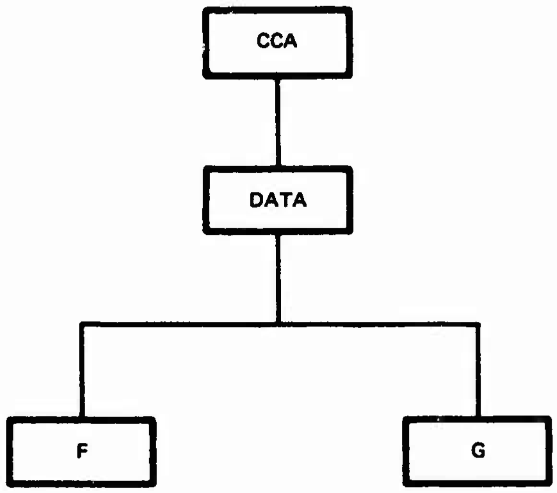

Figure 2-2. A portion of the directory.

Only a bottom-most node of the directory may be a container ident; only an outermost container has its ident entered in the directory.

Normally, the first thing a user does after attaching to the datacomputer is log in to a directory node. For most purposes, he only sees his login node and the part of the directory that is subordinate to his login node. (The LOGIN request is discussed in detail in Chapter 4.)

## Pathnames

Pathnames are used to reference nodes in the directory tree by describing a path through it. They have the general hierarchical form

```
NODE1.NODE2...NODEn
```

where NODE2 is a node directly subordinate to NODE1.

There are several varieties of pathnames. The two classes of directory objects referenced by pathnames are closed nodes (including all nodes that are not outermost containers and all outermost containers that are not OPEN) and OPEN outermost containers. There are three areas in which names can be found: the TOP, LOGIN, and OPEN contexts. Thus there are six possible pathname types, only five of which are reasonable. (A closed node in the OPEN context isn't.)

Closed nodes can be referenced either by a complete pathname (started with the reserved word %TOP), which causes the name search to be anchored at the top of the directory tree, or a LOGIN pathname, which anchors the search at the current LOGIN node. Either pathname may contain passwords. (Passwords are discussed in Chapter 4.)

OPEN nodes may be referenced by a simple complete pathname or a simple LOGIN pathname, neither one of which can contain passwords, or by an OPEN node simple name. An OPEN node simple name is the name of the outermost container.

## Creating Nodes

A node in the directory is created with a CREATE request. Such a request has the form

```
CREATE <pathname> ;
```

Only one node may be created by a single CREATE request, and a higher-level node must always be created before one subordinate to it. The reserved words listed in Appendix B may not be used as directory node names.

As an example, let us create the outermost container F, a LIST of 4-character strings; the container's ident will be entered in the directory as indicated in Figure 2-2. We assume that nothing is presently in the directory, so we must start by creating the topmost node.

```
        CREATE CCA;
        CREATE CCA.DATA;
        CREATE CCA.DATA.F FILE LIST
                    FOO STR (4) ;
```

Now that CCA and CCA.DATA have been created, we could create CCA.DATA.G with only one CREATE request; i.e.

```
        CREATE CCA.DATA.G PORT LIST  etc.
```

## Creating Containers

Outermost containers are created by a more complicated form of the CREATE request. The CREATE statement must tell the datacomputer all about the container, for example, its ident, function, size, and data type are included. An outermost container and all its subcontainers must be created at once, with one CREATE request.

The CREATE request causes the description to be stored. It also causes space to be allocated if the container is a FILE.

The full BNF in Appendix A indicates succinctly the precise syntax of the CREATE statement. It is worth looking at a few examples before looking at all the details of descriptions. A LIST of STRings:

```
        CREATE ALPHA FILE LIST SUBCONTAINEDSTRING STR (44) ;
```

Here the size of the outermost LIST is omitted, so the datacomputer will calculate a default size.

A LIST of STRUCTs, each of which contains three strings:

```
     CREATE BALLTEAM FILE LIST (25)
                PLAYER STRUCT
                  NAME STR(20)
                  POSITION STR(2)
                  UNIFORM%NUMBER STR(2)
                END;
```

The datacomputer will allocate enough space for the file BALLTEAM to hold 25 copies of the STRUCT named PLAYER. Note that END is required to terminate the description of the STRUCT.

The example diagrammed on page 4:

```
       CREATE PEOPLE FILE LIST
          PERSON STRUCT
```

```
         NAME STRUCT
           FIRST STR(15)
           LAST  STR(15)
         END
         ADDRESS STRUCT
           STREET STR(15)
           CITY STR(15)
           STATE STR (15)
         END
         SOCSECNO STR(10)
       END;
```

The elementary data types are BYTE and STR. Containers of these types contain data, not other containers.

STRings and LISTs must have a size. For STRings, the size is the number of bytes in the STRing. For LISTs the size is the number of LIST members (e.g. the number of PERSONs in PEOPLE above.) The three forms for indicating the size are:

```
(n) -- a fixed size of n
(m,n) -- a minimum size of m and a maximum of n
(,n) -- a minimum dimension of 0 and a maximum of n
```

where m and n are positive integers.

For an outermost LIST or STRing, no size need be specified. The default minimum is 0, and the default maximum is based on what will fit in the default space allocation.

The datacomputer needs a way to find the end of the data in variable-sized LISTs and STRings. The three options are a preceding count, a trailing character, and punctuation (i.e. a device-dependent marker). A one-byte preceding count is indicated with the keyword parameter

```
,C=1
```

Version 0/10 cannot handle counts larger than one byte. Thus, if there is a count, then the maximum dimension must be small enough to fit into a one-byte count. (Byte size is discussed further below.) The value of the count does not include the count byte itself.

The syntax to indicate that there is a one-byte delimiter is

```
,D=n
```

or

```
        ,D='a'
```

where n is a decimal number and a is any ASCII number, letter or special character.

The datacomputer considers punctuation to be different from delimiters. Punctuation over the network is a special character (specifically EOR, EOB, or EOF) inserted in the data but not considered part of the data. This is indicated by

```
        ,P=EOR
        ,P=EOB
and
        ,P=EOF
```

A fixed-size container (including a STRUCT) may have a P, D or C parameter, but no container (fixed or variable) may have more than one of these.

A datacomputer FILE can be punctuated, but none of its sucontainers can be. The FILE punctuation defaults to EOF. Variable-length subcontainers must have either a C (count) or D (delimited) parameter.

If a variable-sized PORT does not have a P, D, or C parameter, then it defaults to P=EOF. Variable-sized subcontainers of a PORT default to P=EOR.

Punctuation is hierarchical. A container that is punctuated with EOR cannot contain one that is punctuated with EOB or EOF. A container that is punctuated with EOB cannot contain one with EOF. If higher punctuation is found in a data stream where the datacomputer is looking for lower punctuation (e.g., an EOB where an EOR is expected), the higher punctuation implies the lower.

- - - - - - - - - - - - - - - - - - - -

(1) Note that the default punctuation for PORTs is different from what it was in Version 0/9. Consider the description

```
        CREATE P PORT LIST
                R STRUCT
                   A STR (1)
                   B STR (1)
                END;
```

FOR VERSION 0/9 EVERY R MUST END WITH AN EOR. In Version 0/10, since R is fixed-size, no EOR's are expected, and an error message is output if an EOR or EOB is found. If R's end with EOR, then

```
        ,P=EOR
```

should be added to the description of R.

The interpretation of a STR is one of ASCII (i.e. 7-bit ASCII), ASCII8, or BYTE, as in the following three examples:

```
A STR ASCII (5)
P STR ASCII8 (1,10), C=1
WALDO STR BYTE (73)
```

The default byte size for BYTE is 36 bits. BYTE is optional if the byte size is given explicitly with the keyword paramter

```
,B=n
```

where n is a positive integer less than or equal to 36. The ,B=n option may not be used for ASCII STRings. If no byte size or interpretation is given, then the STR is 7-bit ASCII.

At times the datacomputer needs to fill in a value or a part of a value. The user can specify a fill character thus:

```
,F='a'
```

or

```
,F=n
```

where a is an ASCII character and n is a decimal number. The default fill character is blank for ASCII data and zero for non-ASCII data.

Note that a byte size and a fill character can apply to a STRUCT or a LIST as well as a STR or a BYTE. Consider the following:

```
CREATE F FILE LIST
         R STRUCT, B=36
            A STR (5)
         END;
```

The byte size of A is 7. A takes up 35 bits. There is one "unused" bit after A before the next R. Thus, R must be filled. Even though the data (i.e. A) is ASCII, R is non-ASCII because it does not have a 7-bit byte size. Hence, the default filler of 0 is used for the bit.

The rules for punctuation, byte size and fillers are simple but not at all intuitive. In general, specifying punctuation rather than relying on defaults helps avoid errors. Also

```
LIST <pathname> %DESC;
```

will output a complete description, including all default lengths, dimensions, punctuation, byte sizes and fillers. (The LIST command is discussed more fully below.) It is often instructive to look closely at the %DESC to see where it is different from what the user expects.

BYTEs and STRings that will frequently be used for retrieval may be inverted. For members of outer LISTs, the option

```
,I=D
```

is used. for members of inner LISTs, the option

```
,I=I
```

is used. Inversions and the difference between outer list members and inner list members is discussed more fully in the section on WITH.

# Chapter 3: Directory Commands

## OPEN

Before data can be input to or read from a FILE or PORT, the container must be open, and a mode must be specified for it. The mode of a FILE or PORT, which is set when the container is opened, determines the legality of various operations on that container.

Mode is one of READ, WRITE, or APPEND. Data can only be transmitted out of a FILE or PORT that is open in READ mode, but either out of or into a FILE or PORT that is open in WRITE or APPEND modes. The difference between WRITE and APPEND lies in their treatment of any data that is already in the container when it is opened. When an assignment is made to a container that was opened in WRITE mode, any data it contained previously is thrown away, but a container opened in APPEND mode has newly-arriving data written after the end of any already-present data, which is thus preserved.

A variation of WRITE and APPEND is WRITE DEFER and APPEND DEFER. When DEFER is indicated as part of the mode, a more efficient technique of updating the inversion is used.

When a FILE or PORT is created, it is opened in WRITE mode. A FILE/PORT that already exists may be opened with an OPEN request:

```
OPEN <pathname> <mode> ;
```

which specifies the name of the container that is to be opened and the mode of opening. The name can be either a complete pathname (started with the reserved word %TOP) or it can be a login pathname, started with a node immediately subordinate to the current login node. The mode must be one for which the user has privileges (see Chapter 4). The mode argument may be left out of an OPEN statement, in which case the container is opened in READ mode if it is a FILE and WRITE mode if it is a PORT. Two outermost containers with the same ident may not be open at the same time.

For example, to read data that was previously stored in CCA.DATA.F, a file, either

```
OPEN CCA.DATA.F;
```

or, if the current login node is CCA,

```
        OPEN DATA.F;
```

WILL OPEN F PREPARATORY TO DATA TRANSFER REQUESTS.

## MODE

The mode of a container that is already open may be changed with the MODE statement:

```
        MODE  <paragraph> <mode> ;
```

The pathname can be a simple complete pathname (i.e. a complete pathname with no passwords), a simple login pathname, or a node name.

## CLOSE

The complement of the OPEN request is the CLOSE request. When you have finished using an open container, close it with

```
        CLOSE <pathname> ;
```

where pathname must be the simple pathname of an open container. Closing a FILE/PORT with a function of TEMPORARY PORT has the effect of deleting its description from the datacomputer.

## DELETE

The ability to delete directory nodes is useful in maintaining a data base at the datacomputer. The DELETE request allows one to delete one or several outermost containers and all the data they contain.

```
        DELETE <pathname> ;
```

causes the node named by <pathname> to be deleted from the directory. The pathname must be the login pathname. Thus, only nodes subordinate to the login node can be deleted. The node cannot have any subordinates.

```
        DELETE <PATHNAME>.** ;
```

deletes the node and all subordinate nodes. If any of the deleted nodes are outermost containers, the container descriptions and any associated data are deleted as well. The DELETE request need not be used on TEMPORARY PORTs, as they are automatically deleted either when they are closed, or at session end.

If the data stored in FILE is to be deleted, but the container description itself retained in storage, the DELETE request cannot be used. Instead, CREATE a port B with a description matching the container A that is to be emptied, and execute the assignment A = B with no data in B. the effect of this assignment is to delete all the data from A.

## LIST

The LIST request is the means by which the user interrogates the datacomputer about his environment. The request has two arguments: the node or nodes which are the object of the inquiry, and the type of information desired.

The first argument consists of a set of nodes in the directory. Possible node sets are: 1) a single node, 2) all nodes directly subordinate to a given node, 3) a node and all its subordinates, and 4) all open files and ports. A single node is specified with a full pathname, which can include passwords and can be anchored at the top node (%TOP). The set of a node's direct subordinates is indicated with either a "*" (the login pathame is implicit) or a full pathname followed by a "*". Either "**" or a full pathname followed by a "**" designates a node and all its subordinates. The set of all open nodes is referenced by %OPEN. %TOP alone defaults to %TOP.**.

There are five kinds of available information. These are: 1) node names and related data (node type, privileges, and possibly mode and connected argument), 2) parsed data descriptions (of FILEs and PORTs), 3) original source text of data descriptions, 4) allocated space (for FILEs), and 5) privilege blocks associated with nodes. These information options are specified by %NAME, %DESC or %DESCRIPTION, %SOURCE, %ALLOC or %ALLOCATION, and %PRIV or %PRIVILEGE, respectively. The default option is %NAME.

Not all of the kinds of information are available for all of the possible node sets. The options that are available are:

| Node Set | Option |
|---|---|
| <pathname> | %DESC |
| <pathname> | %NAME |
| <pathname> | %SOURCE |
| <pathname> | %ALLOCATION |
| <pathname> | %PRIVILEGE |
| <pathname>.* | %NAME |
| <pathname>.** | %NAME |

```
<pathname>.**             %SOURCE
%OPEN                     %NAME
%OPEN                     %DESCRIPTION
%OPEN                     %SOURCE
%OPEN                     %ALLOCATION
```

# Chapter 4: Security and Passwords

## Introductory Concepts

The 0/10 version of the datacomputer provides file-level security (restricted access to nodes and attendant data) by means of a system of privilege blocks, described in the following sections. One or more (or no) blocks may be associated with a particular node. Each privilege block defines a class of users who may be given access to the node and the set of privileges to be granted to such users. Whenever a user attempts to access a node or file, the datacomputer will scan that node or file's privilege block(s), if any, to ensure that the user is 'legal' and to determine what privileges will be allowed.

## Chapter Organization

This chapter is divided into three principal parts. The first sections describe what privilege blocks are and how they provide file security functions for datacomputer users, and introduce the reader to the security features of datalanguage. The second part completely specifies the datalanguage needed for creating, deleting and manipulating privilege blocks, and completes the description of their components begun in the first part. The third section offers several examples of how to add, delete and look at privilege blocks.

## Gaining Access to Nodes: LOGIN

Every node in the directory has certain privileges associated with it. For example, the ability to create inferior nodes, or to read or write file data, are privileges which may be granted or denied to a particular node. When a user initially connects to the datacomputer he is automatically connected to the top node of the directory tree (%TOP), and he (i.e., the %TOP node) is granted minimal privileges. To acquire more, he must log in to some node, which is called, curiously enough, the login node.

Logging into this node establishes the user's identity for subsequent pathname references (1). It should be kept in

- - - - - - - - - - - - - - - - - - - -

(1) In addition to establishing a user identity for privilege purposes, logging in performs various accounting and pathname context functions.

mind that a user is identified to the datacomputer _only_ by his login node. Thus, throughout this chapter, the terms 'user-id' or 'user name' are to be understood to mean nothing more than the full pathname, including the specified privilege block (if any) at each level (2), of the node to which the user has logged-in.

Whenever a logged-in user references a node, the login pathname is compared against the user-id field of every block in the node's privilege block list. If a block is found whose user class description includes the pathname of the login node, the privilege-set described by the block will be added to (or taken away from) the privilege set already given to the login node.

## Privileges

Privilege set specifications come in two flavors: privileges to be granted (added) to the node and privileges to be denied (taken away). If a privilege is not specified (as either grant or deny), then that privilege (or denial of it) is passed, unchanged, from the superior node to its subordinate. At each node level, the deny bits specified in the given privilege block are NOT-AND'ed with the privileges of the superior node. Then the grant privileges are OR'ed with the result, to yield the privilege set for that node.

It is important to understand that _privileges may be added and taken away at every level of the pathname._ For example, suppose the login node has the privilege set <CLWA> (3), and a subnode's privilege block specifies: grant read privilege (G=R), and deny write privilege (D=W). The result at the subnode would be the final privilege set of <CRA> (4).

(2) Pathnames may be _qualified_ or _unqualified_. A qualified pathname is one containing password strings for the purpose of gaining particular privileges upon opening the node, e.g.,

```
        NODE1('PASSWORD1').NODE2.NODE3('PW3')
```

is a pathname qualified at the first and third levels by the passwords 'PASSWORD1' and 'PW3', respectively. The pathname NODE1.NODE2.NODE3, on the other hand, is unqualified. Prior to Version 0/10, all pathnames were unqualified.

- - - - - - - - - - - - - - - - - - - - - - - - -

(3) This is a shorthand way of saying 'this node has been granted control <C>, login <L>, write-to-file <W> and append-to-file <A> privileges. Specific privileges are described in detail below.

Note that a node can never look at, modify, or affect a superior node in any way not possible at the level of the superior. That is, if a user cannot look at the privilege blocks for a node, he cannot acquire that privilege for that node from an inferior one. However, an inferior node may well have privileges relative to its subnodes that its superior does not have relative to its subnodes. For example, scanning along the pathname A.B.C.D.E...., A.B.C may have only read privileges, but does not have write privilege. Now, the node A.B.C.D may be granted write privilege at level D (thus awarding A.B.C.D read/write privileges), this does not affect A.B.C. It still has only read privilege.

## Privilege Block

Privilege blocks are data structures which define access to nodes. Each privilege block is associated with one particular node. Any node in the directory, including ports and file, may have privilege blocks defined for them. A node may have any number (including zero) of privilege blocks. When an attempt is made to access a node which has privilege block(s), those blocks are scanned for a user-id corresponding to the current login pathname and for a password string matching that supplied by the user in the request referencing the node (e.g., LOGIN, OPEN, DELETE, etc.). If a match is found, the matching block's privilege set bits are examined and the appropriate privileges are granted/denied the node. The matching algorithm is described below in more detail.

Each privilege block can contain:
- user name
- host name
- socket number
- password character string
- grant privileges
- deny privileges

Each of the above fields falls into one of two categories: 1) a description of the group of users which may access the associated node; and 2) the privileges to be granted to these users.

The privilege block is completely specified at the time it is created. When a node is referenced, only the password string, if any, is required; the user-id (including host name and socket number), has been retained by the login process.

(4) The login privilege is not propagated to subnodes. It applies only to the node for which it is explicitly granted. See below.

Privilege blocks are created by the datalanguage command CREATEP. They are deleted by the command DELETEP. Existing privilege blocks may be displayed via the LIST nodename %PRIV(ILEGE) command. The full syntax of these commands is described below.

## User Identification Fields (User-ID)

The user identification fields include some or all of the following: a valid login pathname or a class of login pathnames, the number of a host computer, the datacomputer socket number, and a password character string. These fields are discussed in more detail in the following sections.

## Host

The host name is an optional field. If specified, it must be a decimal number from 1 to 255 designating the number of the host computer. The host name cannot be a number greater than 255, or less than 1. It cannot be a character string, except for the special cases LOCAL and ANY.

LOCAL host indicates that the user should not have connected to the datacomputer via the ARPANET. Effectively, this means (at this time) that the user is located at CCA and is connected to the datacomputer via a local terminal.

The host name may also be ANY, which means that any host, foreign or local, is acceptable.

If a host name is not specified, the default value is ANY.

## User Name

The user name is the pathname or classname (5) of the login node(s) which may gain access to the node associated with the privilege block. Note that a different privilege block must be created for each specific user permitted to use a given password. For example, if two different users, say CCA.WALDO and CCA.DINGLE, wanted to use the same password string ('FOO') to gain access to a node, two separate blocks would have to be created, one per specific user name. Thus, in this example, one privilege block would contain the information CCA.WALDO ('FOO'); the other,

- - - - - - - - - - - - - - - - - - - - - - - -

(5) User classnames are defined below.

CCA.DINGLE ('FOO').

If no user name is specified, the default is **, which grants any user access to the node.

## Socket

The socket number is a 32-bit number, e.g., 600403, or ANY. This is an identification number assigned by the foreign host to the user logged in on that foreign host. Usage of the socket number in the CREATEP statement ensures that only specified users at the foreign host site may gain access to a particular node.

Socket number defaults to ANY.

## Password

A password consists of an alphameric string enclosed by single quote (') characters, e.g., P='FOO'. Non-printing characters, except blanks, are not valid in a password string. Blanks may appear at any point in the quoted string. Tab characters are not permitted.

A privilege block need not contain a password. In this case, none should be given when referencing the node. Note that no password is not the same as, and is is treated differently from, a null password (''). Null password is treated as a password of zero length, and must be supplied as such whenever the node is referenced.

## Privilege set specifications

The following privilege bits are defined for 0/10:

LOGIN (L)

In order to control login identities more closely, the ability to log in to a node is not passed to subordinates. As a result, -L (deny login) is meaningless.

CONTROL (C)

Control includes complete subordinate control and privilege control. Control is required for creating and deleting nodes, file
s and privilege blocks. It
is also required for listing privilege blocks. It is very powerful, and cannot be removed by an inferior: -C is not permitted. After 0/10, C may be split into meaningful components.

```
Data Control Privileges
     READ (R)
     WRITE (W)
          W implies R and A.
     APPEND (A)
          A does not imply R.
     Conflicts are not allowed in one tuple, e.g.
          +R and -R.
```

## Ordering of Privilege Blocks

The ordering of privilege blocks is important. When a node is referenced, the privilege blocks (if any) for that node are scanned linearly for a password string matching the password entered by the user. If a match is found, the user-id of the privilege block is compared to the login identity. If they match, the associated privileges are granted/denied, and access appropriate to the granted privilege set are awarded to the node. If the end of the privilege blocks is reached without finding a password/user-id match, the node is opened with no privileges.

Since the privilege blocks are scanned linearly, their ordering defines their selectivity. For example, suppose a node to have two privilege blocks which specify the same password ('foo') but different login nodes, say, A and **, and suppose that the block with user name A grants greater privileges (read/write/append) than that with ** (which permits read). The proper ordering, as displayed by a

```
LIST WALDO.NODENAME %PRIV(ILEGE);
```

statement, is as follows: (note 6)

```
(1),U=A,H=ANY,S=ANY,G=RWA
(2),U=**,H=ANY,S=ANY,G=R   (note 7)
```

If the order of these blocks were reversed, so that the block with the user name '**' were first, then whenever the password FOO was encountered the first block would be selected; i.e., every login pathname would match the '**', and the matching process would be complete. Thus, the block

- - - - - - - - - - - - - - - - - -

(6) Details of this command are given below.

(7) U=** means that any user name will be accepted as valid.

with the user name A would never be found, and the user A would be unable to open the node with the greater privileges which should be granted him.

In 0/10 the user is responsible for maintaining the desired search order, by adding and deleting privilege blocks via their block index numbers. The datalanguage for this process is described below. Future versions of the datacomputer may provide an automatic ordering algorithm, which could be manually overridden, if desired.

## User Classes ('Star' Feature)

Classes of users may be given access to a node by specifying a user class as the user name instead of a single user. This is done by means of the '*' and '**' ('star' and 'star-star') features. If a star appears in a pathname, it is interpreted to mean: 'any single (non-null) partial pathname is acceptable here'. That is, if the nodes A.B.N1, A.B.N2, and A.B.N3 exist in the directory tree, usage of the user classname A.B.* would specify any of these three pathnames. Stars may appear at any number of levels; for example, if the nodes A.X.N1 and A.Y.N4 exist, then the user-name A.*.* would specify both of these nodes, as well as any of the previous three. The use of a star at any level implies that there must be a partial pathname at that level; e.g., the classname A.*.* could not specify node A or A.J.

## User Classes, cont. ('Star-star' Feature)

The use of a single star in a pathname indicates that a node must exist at the level corresponding to that of the star, and a star must be explicitly specified for each desired level. The star-star feature is designed to permit access to several levels of nodes. A star-star ('**') in a user name is interpreted to mean: 'any number (including zero) of partial pathnames are acceptable here'. Thus, referring to the example of the preceeding paragraph, A.B.N1 could be specified by any of the following:

```
A.B.N1.**
A.B.*.**
A.B.**
A.*.**
A.**
*.**
**
```

For 0/10, only trailing *'s and/or a final ** are allowed. The following, for example, are illegal:

```
A.*.C
```

```
A.**.C
A.*.**.D
A.**.*
*.B.**
**.*
```

## Datalanguage for File Security

Two new datalanguage statements, CREATEP and DELETEP, create and delete privilege blocks. They are discussed in the following sections. The list command has a new option, %PRIV (or %PRIVILEGE), which allows the user to list the privilege blocks for a node.

CREATEP and DELETEP are privileged requests. They are only accepted when the associated node can be referenced with control privilege <C>. (This means that it may be necessary to login to some particular node before any privilege blocks can be added to another, and that passwords may be required for the login process or for referencing nodes superior to the node for which the privilege block is to be added.)

## Creating Privilege Blocks: CREATEP

Privilege blocks are created, and fully specified by, the CREATEP command. A fully specified CREATEP statement might appear as follows:

```
        CREATEP NODE1('PW1').NODE2, U=CCA.WALDO.*.**,  H=34,
S=604320,
                   P='SECRET PASSWORD', G=R, D=WA, N=2;
```

In this example, the node for which we are creating a privilege block is NODE1.NODE2. We must specify ('PW1') for NODE1 in order, perhaps, to gain control privileges at the first level. The parameters which follow the nodename is the privilege keyword list. These are discussed individually in the following sections, and are summarized in Appendices A and B.

## CREATEP: User Name

The user name is specified by 'U=' followed by an unqualified pathname or classname string. The pathname may have any number of levels. It must not contain password strings for any level.

The following are valid pathnames/classnames.

```
CCA
CCA.WALDO.DINGLE
CCA.*.*
CCA.**
*.*.*
*.**
**
```

## CREATEP: Host Number

The host number is specified by 'H=' followed by a decimal number from 1 to 255, or either of the strings LOCAL or ANY.

```
H=28
H=ANY
H=LOCAL
```

## CREATEP: Socket Number

The socket number is specified by 'S=' followed by the 32-bit foreign-host assigned decimal number corresponding to the directory the user is logged into at that foreign host, or the string ANY.

```
S=309483
S=ANY
```

## CREATEP: Password String

The password string is specified by 'P=' followed by any datacomputer string constant (tabs may not be included, although blanks are permitted), e.g., 'PASSWORD 1', '? * ++!!', or '' (null password).

Note that if no password string is specified at CREATEP time, then that privilege block will have no password associated with it. No password is different from null password (P=''), which is a valid password zero characters in length.

## CREATEP: Grant Privileges

Privileges are granted by 'G=' followed by

| | |
|---|---|
| C | (control) |
| L | (login) |
| R | (read file data) |
| W | (write file data) |
| A | (append data to file) |

in any combination and in any order, e.g., G=CRAWL (all privileges), G=WAR (read/write/append), etc.

## CREATEP: Deny Privileges

Deny privileges are specified by 'D=' followed by R, W or A. Login (L) applies only to the node for which it is specified. It is not passed to subordinates. Control (C) cannot be removed by any inferior node, i.e., it is passed to all subnodes.

## CREATEP: Privilege Block Index

As privilege blocks are created, they are assigned index numbers by the datacomputer. Block numbers are assigned to privilege blocks sequentially according to their search order. Block numbers can range from one to n, where n is the total number of password blocks in the search sequence. Blocks can be explicitly ordered by the user at CREATEP time by entering 'N=' followed by the number that the newly added block is to have in the search sequence. N must be greater than zero, and not greater than the total number of privilege blocks currently existing for the node. Note that this index is not in any sense a part of the data contained in the privilege block; it is merely the position of the block in the password block list.

An example, if there were three blocks in the privilege block list for a node (NODE1),

```
1  U=AAA
2  U=CCC
3  U=DDD
```

and a new block were to be added between the first and second existing blocks, i.e., so that the new block would then occupy second position, we add a keyword, N=2, to a CREATEP command:

```
CREATEP  NODE1,U=BBB,P='ZOO',N=2;
```

which results in the following privilege block list:

```
1  U=AAA
2  U=BBB
3  U=CCC
4  U=DDD
```

If N had been omitted, the new block would have been added at the end of the list. Note that the indices of the two blocks following the new one have been bumped by one. Similarly, if any block is deleted, the indices of all the following blocks are reduced by one.

## Looking at Privilege Blocks: LIST

In order to permit the user to list privilege block information, the %PRIV (or %PRIVILEGE) option has been added to the datalanguage LIST request. It looks like this:

```
        LIST CCA.WALDO  %PRIV      (or)
        LIST CCA.WALDO  %PRIVILEGE
```

Passwords cannot be listed with the %PRIV option (or in any other way - so don't forget 'em!). Privilege block information is preceeded by the index number of that block. All other information in the privilege block is listed in a format similar to that which might be found in a CREATEP command, e.g, either of the LIST requests above might generate the following output from the datacomputer:

```
        (1),U=CCA.WALDO,H=LOCAL,S=ANY,G=CRAWL
        (2),U=CCA.*.**,H=ANY,S=ANY,G=RWAL
        (3),U=*.**,H=32,S=654364,G=RL,D=WA
```

%PRIV may be used only when the controlling node has control privileges.

## Deleting Privilege Blocks: DELETEP

Privilege blocks may be deleted with DELETEP followed by the index number of the privilege block to be deleted,

```
        DELETEP 3
```

The controlling node must have control privilege.

## Example

This example will create a node which will be the controlling node for all other nodes at site CCA. Presumably, access to this controlling node would be restricted to very few persons at that site; 'super-users', as it were. This could be done by means of a password. In addition, anyone seeking control privileges for CCA might be required to be logged-in to some other (access restricted) node. The person with access to CCA would be responsible for creating subnodes, perhaps one for each programmer permitted to use the datacomputer. These individual programmers could then create their own directory structures (nodes, ports and files) in any manner they wish.

The site-node CCA is created by the following series of requests:

```
CREATE CCA;
CREATEP CCA,P='HONCHO',G=CL;
CREATEP CCA,P='FLUNKY',G=L;
LOGIN CCA('HONCHO');
```

The user is now logged in to CCA. He has control privileges. Next he creates a series of programmer-nodes, each with control privileges. Initially, two privilege blocks are created for each programmer node. One requires a password (known to, and probably specified by, the individual programmer), and the other requires no password and is accessible to anyone logged in to CCA or any of its subnodes. However, persons who log in to a programmer node without specifying a password are not given control privileges and thus cannot modify or delete anything that the programmer wishes to keep secure.

```
CREATE WALDO;   CREATEP WALDO,U=CCA,P='TURKEY',G=CL;
                CREATEP WALDO,U=CCA.**,G=L;
CREATE CLYDE;   CREATEP CLYDE,U=CCA,P='FETCH',G=CL;
                CREATEP CLYDE,U=CCA.**,G=L;
      .
      .
      .
      .
CREATE DINK;    CREATEP DINK,U=CCA,P='PODUNK',G=CL;
                CREATEP DINK,U=CCA.**,G=L;
```

After this is done, super-user checks the privilege blocks he has created, first at his own node level:

```
LIST %TOP.CCA('HONCHO') %PRIVILEGE;
```

and he receives a datacomputer printout in the following format:

```
(1),U=**,H=ANY,S=ANY,G=CL
(2),U=**,H=ANY,S=ANY,G=L
```

He next verifies that each of the programmer-node privilege blocks has been correctly entered, e.g.,

```
LIST WALDO %PRIV;
```

and the datacomputer replies:

```
(1),U=CCA,H=ANY,S=ANY,G=CL
(2),U=CCA.**,H=ANY,S=ANY,G=L
```

At this point, programmer Waldo tells super-user that he would rather have 'donkey' as his control password rather than 'turkey'. Since the user name (U=CCA) in Waldo's control privilege block is more restrictive than the user name (U=CCA.**) in the non-control privilege block, the first privilege block must be deleted and the new one added in the same position (N=1):

```
DELETE WALDO 1;
CREATEP WALDO,U=CCA,P='DONKEY',G=CL,N=1;
```

We now have the following directory:

```
CCA
CCA.WALDO
CCA.CLYDE
   .
   .
   .
CCA.DINK
```

Each of the programmer-nodes listed above has its own password which is known to the person having access to that node. In addition, each is required to login to CCA before being able to acquire login and control privileges at its own level. (Most or all of the programmers at CCA are given only the password FLUNKY, which does not give control privileges. Thus, they cannot create or delete any nodes at the programmer-node level or look at the restricted data of any other programmers.)

As soon as he is informed that he may join the select international hoard of datacomputer users, Waldo rushes to his terminal to login:

```
LOGIN CCA('FLUNKY');
LOGIN WALDO('DONKEY');
```

Since he has logged in to his node using the password which grants control privileges, Waldo now creates BOOKFILE and BOOKPORT and reads some data into BOOKFILE from a TENEX file named TENEX-BOOK.FILE (note 8):

```
CREATE BOOKFILE FILE LIST(,1000),P=EOF
  BOOK STRUCT
        TITLE STR (,100),C=1
        AUTHORS LIST(,5),C=1
          AUTHOR STR (,50),C=1
        PUBLISHER STR (,50),C=1
  END;

CREATE BOOKPORT PORT LIST(,1000),P=EOF
  BOOK STRUCT
        TITLE STR (,100),P=EOR
        AUTHORS LIST(,5),P=EOB
          AUTHOR STR (,50),P=EOR
        PUBLISHER STR (,50),P=EOR
  END;

CLOSE %OPEN;

OPEN BOOKFILE WRITE;
OPEN BOOKPORT; CONNECT BOOKPORT 'TENEX-BOOK.FILE';
 (NOTE 8)
BOOKFILE=BOOKPORT;

CLOSE %OPEN;
```

In order to permit others to look at his library file, Waldo creates a couple of privilege blocks. The first permits anyone at CCA to look at his book list, while denying him the right to change anything. The second is for Waldo's private use in changing the file:

```
CREATEP BOOKFILE,U=CCA.*,G=R,D=AW;
CREATEP
 BOOKFILE,U=CCA.WALDO,P='READ*MORE*EVERY*DAY',G=RWA;
```

- - - - - - - - - - - - - - - - - -

(8) A TENEX filename is used in this example for the purpose of didactic clarity. In practice, this would usually be done only by local datacomputer users (users located at the site of the datacomputer). Remote users would have to arrange for operator intervention, if connecting to a file at the datacomputer site; or would specify the host name and socket number from which the data would be sent to the datacomputer.

# Chapter 5: Assignment and For-loops

## Assignment Involving Outermost Containers

Transmission of data is achieved with an assignment. The syntax of an assignment request that involves two outermost containers is

```
<ident> = <ident>;
```

where the <ident>s are the node names of open outermost containers. The first ident in the statement is that of the receiving container; it must be open in either WRITE or APPEND mode. The second ident is that of the transmitting container; it can be open in any mode, but it must have READ privilege (see Chapter 4).If the second ident is a FILE, it must contain some data.

The containers in the assignment may be either files or ports. The various combinations are listed here, with a description of the action of the assignment request in each case.

| Receiving container | Transmitting container | Comment |
|---|---|---|
| FILE | FILE | copies data from one FILE to another within the datacomputer. |
| FILE | PORT | transmits data from some source external to the datacomputer through a PORT, into a FILE. |
| PORT | FILE | transmits data from a FILE, where it is being kept in the datacomputer, through a PORT, to the outside world. |
| PORT | PORT | transmits data from one place to another in the outside world, using the datacomputer only as a channel for transmission. |

## The Matching Rules

In any assignment statement such as

```
X = Y;
```

(not only one involving two outermost containers) the two operands, X and Y, each has its own description. The datacomputer will transform the data in Y to match the description of X. In order for the datacomputer to be able to do this, the descriptions must _match_. This amounts to a restriction that only similar objects can be assigned to each other. Specifically, for two assignment-operands X and Y to match:

1.A. X and Y must have the same _type_: LIST, STRUCT, or STR, or BYTE,

AND

1.B. If X and Y are both LISTs, then they must have compatible _sizes_, or else X must be a PORT. The sizes are compatible if the minimum size of X is less than or equal to the minimum of Y and the maximum size of X is greater than or equal to the maximum size of Y. This restriction leads to cases where it is legal to assign Y to X but not to assign X to Y. Note that if X and Y are outermost lists with no list size specified, the datacomputer supplies a default size based on the space allocation. (use the LIST request with the %DESC option to find out what the default size is.)

AND

1.C. If X and Y are STRUCTs or LISTs, then at least one container immediately enclosed in X must match, _and have the same ident as_, one container immediately contained in Y,

OR

2. X must be a STRIng and Y a constant. A constant is an arbitrary string of characters. If they are enclosed by single quote marks, then is is an ASCII constant; a single or double quote mark may be included in such a string only by prefixing it with another double quote. The constant 'DON"'T' represents the string DON'T. (This rule is included here for completeness and will be discussed later.)

## _Padding and Truncation_

If two containers of type STR are used in an assignment, the matching rules do not require that their sizes match. There are three cases:

1. The two sizes are equal. The string is assigned without change.

2. In the assignment X=Y, the size of X is greater than that of Y. In this case, it is as if the string in Y is padded at the right-hand side to make it as long as X, before assignment is performed. If a fill character is specified in the description of X (i.e. if the parameter

,F='a' or ,F=n is used in the CREATE request), then that character is used. Otherwise, a blank is used for ASCII strings and zero is used for non-Ascii data.

3. The size of X is less than that of Y. The string contained in Y is truncated at the right-hand side to be as short as X, and the shortened string is then assigned.

## Examples

Let us consider a few examples of the operation of the rules. Suppose we have

```
CREATE M FILE LIST (25), P=EOF RECORD STR(10);
CREATE N TEMP PORT LIST (25), P=EOF RECORD STR(10) ;
M = N;
```

where M is a FILE in which data read from the PORT N is to be stored in the datacomputer. The assignment M = N is legal because M is in WRITE mode and both M and N are open (opened by the CREATE statements). In addition, M and N match: their subcontainers have the same ident (RECORD), and matching descriptions. They satisfy rule 1.A, since the type is STR in both cases, and rules 1.B and 1.C do not apply to containers of type STR.

The effect of this assignment is to read strings of length 10 from the PORT N, and to store them in the FILE M. If an attempt is made to store more than 25 strings in M, the datacomputer will complain, as space was allocated for only 25 strings. However, the 25 in the PORT description is ignored.

A similar example, using the above description for M:

```
    OPEN M APPEND;
    CREATE O TEMP PORT LIST, P=EOF
            RECORD STR (,15), P=EOR ;
    M = O;
```

Each STRing in O is no more than 15 ASCII characters and ends with an EOR. Each one will be padded or truncated to 10 characters since M has fixed-length rather than punctuated STRings.

Now a more complex example.

```
CREATE FF FILE LIST (25), P=EOF
     PERSON STRUCT
        NAME STR (15)
        ADDRESS STR (20)
        CITY STR (10)
        STATE STR (2)
        ZIP STR (5)
```

```
                SOCSECNO STR (10)
                DEPENDENTS LIST (10) NAME STR (15)
             END ;
      ... requests that store data in the FILE FF ...
       CREATE PP PORT LIST, P=EOF
           PERSON STRUCT, P=EOR
               NAME STR (15)
               SOCSECNO STR (10)
           END;
          PP = FF ;
```

Here, the assignment PP = FF is legal because: PP is in WRITE mode, both FF and PP are open, and their descriptions match. Rule 1.A: the type of both FF and PP is LIST. Rule 1.B: PP is a PORT. Rule 1.C: the subcontainer PERSON immediately contained in FF has the same ident as PP.PERSON, and the two STRUCTs PERSON match. We determine this last fact by going round once again with the matching rules.

Rule 1.A: FF.PERSON and PP.PERSON have the same type, STRUCT. Rule 1.B does not apply to STRUCTs. Rule 1.C: a container immediately contained in FF.PERSON, FF.PERSON.NAME, has the same ident (NAME) and a matching description (STR (15)) as a container immediately enclosed by PP.PERSON, that is, PP.PERSON.NAME.

The effect of this assignment is to create a new instance of the struct PP.PERSON for each instance of PERSON in FF, and add it to the LIST PP (that is, output it through the PORT PP). Each PERSON that is output contains only a selection of the data stored in FF: only the NAME and SOCSECNO.

If the situation here were reversed, that is, if FF were open in WRITE mode, and PP were in READ mode, the effect of the assignment

```
          FF = PP;
```

would be to read data from the PORT PP and store it in the FILE FF. However, only the NAME and SOCSECNO would be available as data. The datacomputer handles this situation by assigning strings consisting only of blanks (the default since no fill character is specified in the description) to the unmatched STRs in the output LIST-member. Thus, ADDRESS, CITY, STATE, ZIP, and all 10 instances of DEPENDENTS.NAME would be blank in the FILE FF.

Very often, assignment at the level of outermost containers is all that a user's program will require of the datacomputer. An example would be a time-sharing monitor system, which might want to store backup files, large files, or infrequently-used files at the datacomputer rather than locally on (expensive) disc storage devices. Typically, such a monitor system would itself keep track of where various files resided, and move them from place to place over the ARPA network without burdening its users with the details of exactly where their files were stored.

For such an application, a directory might be set up with one node identifying the operating system that is doing the file storage. Subordinate to this node might be the user idents of its various time-sharing users whose files might be stored on the datacomputer. These user nodes, in turn, would have the file-names themselves as subordinate nodes; as bottom-most nodes, these would also be outermost containers and thus could store the data itself. As a diagram:

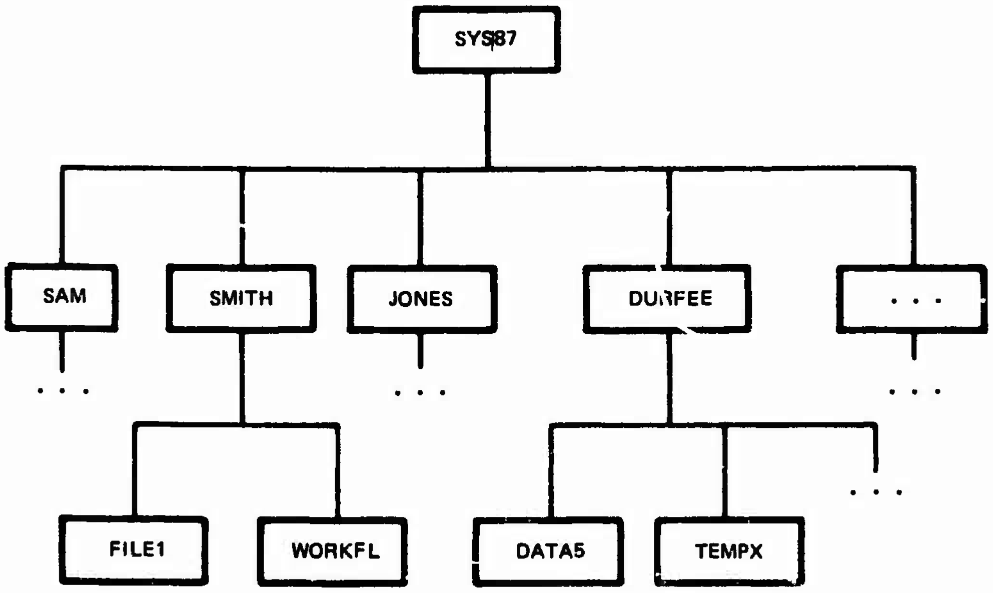

Figure 5-3. The directory for a sample application: providing backup file storage for time-sharing users

A directory of this sort would initially be set up by several CREATE requests; i.e.

```
CREATE SYS87;
  CREATE SYS87.SAM; CREATE SYS87.SMITH;
  CREATE SYS87.JONES; etc.
```

Then, whenever a particular file was to be moved to the datacomputer, a directory node for that file would be set up by, for example,

```
CREATE SYS87.SMITH.FILE1 FILE LIST (999)
            A STR(80);
```

(describing a file with less than 1000 80-character records) and the file would be moved with an assignment statement specifying a PORT with a matching description, and the FILE FILE1, open in WRITE mode. Thus:

```
        CREATE T TEMP PORT LIST A STR(80);
          FILE1 = T;
```

Note that the two outermost containers FILE1 and T in the assignment statement FILE1 = T match each other.

In order to recover the file from the datacomputer when it is again needed, a PORT would be opened in WRITE mode with

```
       CREATE T TEMP PORT LIST A STR(80);
        OPEN SYS87.SMITH.FILE1 READ;
       T = FILE1 ;
```

and the reverse assignment would take place.

## Selection of LIST Members

In the examples given above, there is one output LIST member for every input LIST member. Subsets of the input LIST member (i.e. the LIST on the right side of the =) may be specified by the use of WITH clause as an input-spec. For example, consider the description

```
         CREATE F FILE LIST, P=EOF
                  P STRUCT A STR(3) B STR(5) END;
```

and a matching PORT R. If only some of the P's on the LIST F were to be output -- those with the string A equal to the string '500', say -- one could specify

```
       R = F WITH A EQ '500';
```

referring to the set of all members P of the LIST F that have the given property. Note that A is understood to refer to F.P.A; see the section on the context rules below for an explanation. Quotes are used in the expression '500' to indicate tnat an ASCII string constant is intended.

In a WITH clause, the expressions one can use to choose certain LIST-members, which are called Boolean expressions, must involve comparison of a container that is a STR or BYTE with a constant (like '500' in tie example), using the comparison operators

```
        EQ   (equals)
        NE   (not equal to)
        GT   (greater than)
        LT   (less than)
        GE   (greater than or equal to)
and     LE   (less than or equal to).
```

Combinations of comparisons with

```
OR, AND, NOT, and ANY
```

are also possible. In precedence of operators, ANY (see below) is highest; NOT is next in precedence, then AND, which is in turn higher than OR; parentheses may be used to affect the order of evaluation of these operators.

When using an input-spec, the name of the input LIST-member may be used instead of the name of the input LIST. (This is for consistency with the syntax of the FOR-loop, discussed below.) Thus,

```
R = F.P WITH A EQ '500';
```

is equivalent to the example above. Some sample input-specs are thus:

```
F.P
F.P WITH A EQ '500'
F WITH A EQ '500' AND B GT 'AZZZZ'
F.P WITH (A EQ '500' AND NOT B GT 'MONDA') OR
         (A EQ '600' AND B NE 'ZYYYY')
```

For ASCII containers, the operators GT, LT, etc. compare the ASCII codes for the given strings and the given strings must be of the specified length. This means that the character blank is less than the digits, which in turn are less than the letters. Consult a reference document for the complete list of ASCII codes for all characters.

Also, while an input-spec like

```
F.P WITH A EQ '5'
```

is legal, it will not find any P's, since there are no A's with only one character.

## Retrievals Using Inner List Members

Consider a description like

```
G FILE LIST, P=EOF
  R STRUCT
```

```
                    A STR (4)
                    B STR (4)
                    W LIST (20)
                     WA STR (5)
               END
```

Each R has 20 Wa's, since R contains an inner list (W). An input-spec like

```
        G.R WITH WA EQ 'ABCDE'
```

specifies all R's with at least one Wa with value 'ABCDE'. This may also be expressed as

```
        G.R WITH ANY WA EQ 'ABCDE'
```

The former is called an _implicit_ _ANY_ and the latter, an _explicit_ _ANY_.

The container WA can be used in boolean expressions such as

```
                 G.R WITH
                        ANY (WA EQ 'MARCH' AND WA EQ
 '33103')
                 G.R WITH ANY
                        (WA EQ 'MARCH' OR WA EQ 'WORD ')
                 G.R WITH ANY WA EQ '12345' AND B EQ 'CALI'
```

An ANY expression cannot be used within the object of another ANY expression (nested ANY's).

In most cases, the explicit ANY is not required. However, consider the description:

```
        FAMILIES  FILE LIST (100), P=EOF
          FAMILY STRUCT
                MOTHER STR (10)
                FATHER STR (10)
                CHILDREN LIST (10)
                        CHILD STRUCT
                          NAME STR (,10), C=1
                          AGE STR (2)
                        END
        END;
```

The following expressions are not equivalent:

```
        FAMILY WITH ANY (CHILD.NAME EQ 'ELLEN' AND
           CHILD.AGE EQ '21')
        FAMILY WITH CHILD.NAME EQ 'ELLEN' AND
           CHILD.AGE EQ '21'.
```

The latter case is interpreted as:

```
FAMILY WITH ANY CHILD.NAME EQ 'ELLEN'
   AND ANY CHILD.AGE EQ '21'
```

and refers to any FAMILY with an ELLEN who either is 21 or has a sibling who is 21. The former only refers to FAMILYs with a 21-year-old ELLEN.

In all of these examples, the inner list is the second-level list. If there is a third level list, its members may not be used in a boolean expression. For example, given the description:

```
F FILE LIST R STRUCT
        A STR(1)
        L LIST (5)
                L1 LIST (5)
                        B STR (1)
END;
```

L1 is a third-level list, and so B cannot be used in a WITH expression. However, A may still be used in a WITH expression.

## Retrievals Using Inverted Containers

A STR may be _inverted_ if it is contained in a FILE which is a LIST and if the LIST members are fixed-size. This is useful if the STR will be used often in a boolean expression. Inversion is specified by "I=O" or "I=I" as follows:

```
CREATE F FILE LIST (0,100), P=EOF
  P STRUCT
   A STR (3), I=O
   Q LIST (10)
     B STR (5), I=I
  END;
```

The "I" of the above stands for inversion, the "=O" is used with members of outer lists, the "=I" with inner lists.

An inversion on the string A greatly increases the efficiency of retrieving sets of outermost-LIST members by the contents of the string A -- that is, retrieving subsets of the P's that are defined by their values of A. Retrieval by content based on a particular string is _possible_ whether or not that string is inverted; only the efficiency is improved by the existence of an inversion on the string.

There is a certain cost associated with inversion, however. storage space must be allocated for a secondary data structure that the datacomputer uses for retrievals based on inverted strings. Updating a FILE takes longer when it is inverted, since the secondary data structure must be updated as well. Thus, the decision to invert a particular string will depend on the relative cost of increased retrieval time versus increased storage space, the frequency of retrieval based on the particular string, and other considerations. Appendix C contains further technical details concerning inversion.

## Assignment with FOR

Containers that are not outermost can also be used in assignment statements. With FOR, assignments that retrieve subsets of LIST-members may be performed, in contrast with assignment of outermost containers. FOR causes some set of datalanguage statements (usually assignment statements) to be executed several times, once for each member of a given set of LIST-members.

The syntax of the FOR-request is:

```
FOR <output-spec>, <input-spec> <body> END ;
```

The <input-spec> specifies a set of LIST-members to which the operations specified in the <body> are to be applied. A new member of the LIST specified by the <output-spec> is created for each member of the input set processed. If the output-spec is omitted, the FOR-request generates no output.

The input-spec The input-spec must specify a set of LIST-members. The simplest kind of input-spec is just an entire LIST -- i.e. the set of all the LIST-members. However, the name of the LIST-member and not the LIST itself must be given. For example, if

```
CREATE F FILE LIST, P=EOF
              P STRUCT  A STR (3)  B STR (5)  END;
```

then F.P would be a legal input-spec, and would refer to the set of all P's in the LIST F.

A subset of the LIST-members may be specified by the use of a WITH clause in the input-spec. The input-spec on a FOR-loop looks like the input spec on the assignment of outermost containers (discussed above), except that the LIST-member must be named rather than the LIST. Thus

```
F.P WITH A EQ '500'
```

can be used in a FOR-loop, but not

```
F WITH A EQ '500'
```

_The output-spec_ The output-spec is an optional argument. Like the input-spec, it must be the name of a LIST-_member_. The LIST that contains the LIST-member specified by the output-spec is often called the _output LIST_. A new member is created and added to the output LIST for each execution of the FOR-body.

A FOR-loop may be loosely thought of as assignment between two LISTs. However, the descriptions of the members of the input and output LISTs need not match. Otherwise, the restrictions governing the input and output LISTs of a FOR are largely the same as those governing outermost LISTs used in assignment:

1. Both LISTs must be open or contained in open outermost containers.
2. If the output LIST is an open outermost container, it must be in WRITE or APPEND mode.
3. If the input LIST is not an outermost container, the LIST that most immediately encloses it must be the input LIST of an enclosing FOR loop.
4. Similarly, if the output LIST is not outermost, the LIST that most immediately encloses it must be the output LIST of an enclosing FOR.

_The FOR-body_ The operations that are legal in a FOR-body are assignment and another (nested) FOR. The assignment may be of the form

```
<name> = <constant> ;
```

where <name> refers to a container that is a STR (see matching rule number 3), or assignment may be of the form

```
<name> = <name>;
```

to transfer data from one container to another. If the latter is the case, then assignment is subject to

1. the restrictions specified in the matching rules above,
2. the usual restriction that data can be transmitted into a container only if it is open in WRITE or APPEND mode, and
3. the restriction that assignment must occur between _objects_, not _sets of objects_.
4. In Version 0/10 of datalanguage, there are other restrictions governing the containers that can be referenced in the body of a FOR-loop. See Appendix E.

Let us look at a few examples, and describe their operation in words. With F a FILE as above, and

```
        CREATE Q FILE LIST
            P STRUCT
                A STR (3)
                B STR (5)
            END;
        ...
        OPEN F WRITE;
```

then

```
        F = Q;
```

and

```
        FOR F.P, Q.P
            F.P = Q.P ;
        END;
```

have the same effect: a new member P is created and added to the LIST F.

Likewise

```
        FOR F.P, Q.P WITH A EQ '500'
             F.P = Q.P;
          END;
```

HAS THE SAME EFFECT AS

```
        F = Q.P WITH A EQ '500'
```

A final example: with FF.PERSON and PP.PERSON as given in the example for the matching rules,

```
       FOR PP.PERSON, FF.PERSON WITH STATE EQ 'RI'
                   OR STATE EQ 'CT' OR STATE EQ 'MA'
                   OR STATE EQ 'VT' OR STATE EQ 'NH'
                   OR STATE EQ 'ME'
             PP.PERSON.NAME = FF.PERSON.NAME;
       END;
```

will have the effect of outputting through the PORT PP, the NAMEs of all PERSONs in the FILE FF who live in New England; i.e. with STATE equal to one of the New England states.

# Chapter 6: Using the Datacomputer

We proceed now from the basics of the language itself, such as containers and assignment, to a broader view of how datalanguage might be employed by a user's program. We will discuss such matters as accessing the datacomputer, transmitting data to and from datalanguage PORTs, and various aids to the maintenance of data and FILE and PORT descriptions on the datacomputer.

## Interacting with the Datacomputer

Typically, datalanguage requests will be sent to the datacomputer by a user program residing on some computer on the ARPA network. All interaction between the user program and the datacomputer takes place over the network.

Information transmission over the network takes place along uni-directional paths. For a two-way conversation, two such paths are needed, one for transmission in each direction. The end of a transmission path is called a _socket_; a socket can be either a send (output) or receive (input) socket. Obviously, a transmission path requires a send socket at one end and a receive socket at the other. A diagram of the sockets involved in a two-way conversation over the network appears below.

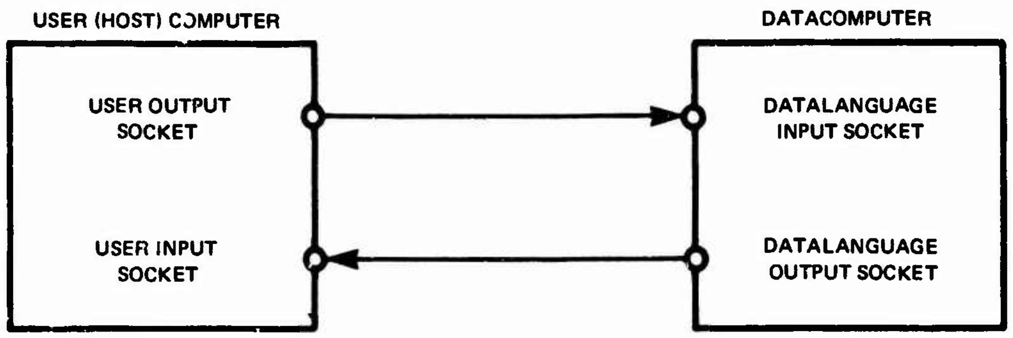

Figure 4-1. Network connections to the datacomputer

A host computer is identified on the network either by a number or by an alphabetic name, like BBN-TENEX. A socket within a given host is identified by a number; send sockets have odd numbers and receive sockets even ones. For a connection to be opened, both hosts involved must request that it be opened. Likewise, after data transmission is complete, both hosts must close their ends of the connection. The period of time during which network connections are open between a user host and the datacomputer is called a <u>session</u>.

In the user program's dialogue with the datacomputer, the transmission in one direction consists largely of datalanguage requests, while messages from the datacomputer are sent in the other direction, to the user program. The sockets at the datacomputer that are used for these purposes are called the <u>datalanguage input socket</u> and the <u>datalanguage output socket</u>. The terms datalanguage input/output <u>port</u> are also used. These ports, like the PORTs that a user can create with datalanguage CREATE requests, are channels for the input and output of information. However, the purpose of the datalanguage ports is to receive datalanguage and transmit datacomputer messages; the purpose of a user PORT is to transmit or receive data.

The protocol by which a user program can set up datalanguage input and output sockets connected to its own output and input sockets is described in Appendix D of this document.

## Synchronization

Since use of the datacomputer typically involves the interaction of two programs at opposite ends of a communication network with a finite time delay, steps must be taken to ensure that the programs remain in synchrony with each other. If they do not, the user program might blithely go on sending datalanguage when the datacomputer expects data or might receive diagnostic messages when it expects a list of directory node names.

To avoid such problems, the datacomputer generates a variety of messages that keep the user program informed of what is going on. The messages fall into several categories: there are error messages, which will be discussed in a later section; informational messages, which can safely be ignored or merely logged by a user program; and synchronization messages, some of which at least must be processed by the user program to ensure proper communication. The first character of the message differs from category to category, allowing the user program easily to differentiate the various classes of message.

| Prefix | Type of Message |
|---|---|
| ?, -, or + | error message |
| ; | informational message |
| . | synchronization message |

Other special characters may be added as datacomputer message-prefix characters in future versions. The letters, digits, tab, and space will never be used as message prefixes, however.

The datacomputer's messages all follow a common format, which includes the special header character just described, a letter and three digits that a program can use to identify the message, the date and time of the message's transmission, and a variable-length string of text that can be read by a human user. Specifically, the format is:

```
.X999 dd-mm-yy hhmm:ss (TAB) TEXT STRING (CR, LF)
```

where . represents the header character, X999 represents the message identifier (for example, I210), dd-mm-yy represents the day, month, and year (for example, 25-09-73), hhmm:ss represents the time on a 24-hour clock in hours, minutes, and seconds, (TAB) represents a tab character, and (CR, LF) represents the carriage return, line feed characters that terminate the message. All alphabetic characters in the message are capitalized. Note that the message may be very long (too long to print on a 72-column printer, for instance), so a user program that processes datacomputer messages may have to format them to be readable.

In this manual, only the invariant parts of messages will be displayed; that is, the header character, the identifying letter and digits, and the message text.

To illustrate the use of synchronization messages in pacing interaction with the datacomputer, consider these two:

```
.I210 LAGC: READING NEW DL BUFFER
.J900 FCFINI: END OF SESSION
```

The first message, .I210, is sent by the datacomputer over the datalanguage output socket, and hopefully received by the user program over an input socket, whenever the datacomputer is ready to accept datalanguage requests. The user program will in general respond to this message by transmitting a _line_ of datalanguage. A line is some number of characters (currently there is an upper limit of about 2500) terminated by either the character sequence carriage return, line feed (ASCII codes 15, 12 octal) or the single character eol (37 octal). On a line may be one datalanguage request (terminated by a semicolon), several requests (each terminated by a semicolon), or a portion of a request.

In the first two cases, when the datacomputer receives the requests (and if they contain no errors) it will proceed to execute them, (typically generating messages and/or initiating data transfers as it does). Following execution, it will again send the .I210 message signifying that it is again ready to receive datalanguage. In the third case, the datacomputer will continue to send .I210 messages, prompting the user program for lines of datalanguage, until a complete request has been assembled; the request will then be executed as described above.

The second message, .J900, is sent by the datacomputer at the end of a session. The user program may request that the session end by sending the datacomputer a control-Z (ASCII code 32 octal) in response to a .I210 message. The datacomputer responds to control-Z by executing an end of session procedure, which involves closing any open containers, deleting TEMP PORTs, and sending the .J900 message. The user program may then close its network connections with the datacomputer.

Synchronization after an error is discussed in the section entitled Error Messages below.

## Transmitting Data through the Datalanguage Ports

Often, a user program will need to send data over the network to be stored at the datacomputer, or to process data that it receives from the datacomputer. If all of the data is described as ASCII, then this may be done by using the datalanguage input or output port.

To reference data that he or she will transmit through the datalanguage input socket, the user need only open a PORT and use it on the right-hand side of an assignment in datalanguage. When the assignment is executed, data will be accepted through the datalanguage input port and assigned to whatever container appears on the left side of the request.

Similarly, to output data through the datalanguage output socket so that it can be picked up by the user program, all that is needed in datalanguage is a PORT used on the left-hand side of an assignment. Any data assigned to that container will be transmitted out through the datalanguage output port over the network.

Of course, performing this feat requires the use of more synchronization messages. To treat the data-input case first:

```
.I231 OCPBO: (DEFAULT) INPUT PORT OPENED
.I251 OCPBC: (DEFAULT) INPUT PORT CLOSED
```

After the user program has sent the datalanguage assignment request that references the open input PORT, the datacomputer will transmit the .I231 message over the datalanguage output port. The message signals that input data is now expected through the datalanguage input port, and the user program should send the data. Data transmission is terminated by a control-Z character, which causes the datacomputer to send the .I251 message confirming that data transmission is finished. The next synchronization message will be .I210, a request for more datalanguage.

The synchronization procedure governing data output through the datalanguage output port is similar. The messages are

```
.I241 OCSOP: (DEFAULT) OUTPUT PORT OPENED
.I261 OCSCL: (DEFAULT) OUTPUT PORT CLOSED
```

When the assignment statement is executed which requests that data be output through the datalanguage port, the datacomputer first sends .I241, followed by the requested data. followed in turn by .I261. The datacomputer does not output a control-Z at the end of the data. The user program can use these messages to separate out the data from all other information.

## Opening a Secondary Port

Instead of a datalanguage port, an additional network connection or secondary port can be used for transmitting data. Non-ASCII data, including an ASCII STR with a preceding count or a non-ASCII delimiter, must be transmitted over a secondary port. The CONNECT request sets up the secondary port.

The CONNECT request names an open PORT, and gives a host (that is, a computer on the network) and socket number to which that PORT is to refer. As mentioned above, if a CONNECT request is never executed for a PORT, it will refer to the socket from which the user program transmits datalanguage (if it is a READ PORT) or the socket at which the user program receives the datacomputer's messages (a WRITE or APPEND PORT). The form of the CONNECT request is

```
CONNECT <pathname> TO <address> ;
```

where <pathname> is the node name, comlete name (i.e. starting with %top) or simple login name (i.e. starting immediately subordinate to the login node) of an open PORT, and <address> can have several forms. It can be one of

| | |
|---|---|
| <socket-no> | the decimal number of a socket at the user's host computer. |
| <host-no> <socket-no> | where <host-no> is the decimal number of a computer on the ARPA network |
| '<host-name>' <socket-no> | where <host-name> is the host computer's TENEX alphabetic name |
| <host-name> <socket-no> | where <host-name> is the host computer's TENEX alphabetic name (such as 'CCA') |
| OR '<local-file-designator>' | This last form of <address> does not refer to the network, but is included here for completeness. <local-file-designator> is a TENEX file designator that refers to a file at the datacomputer site. |

A CONNECT may be executed any time the PORT is open, but it does not actually establish the network connection. Those connections are established, used, and then closed again during the execution of an assignment statement in datalanguage, and CONNECT merely sets up the socket address to be used when the PORT is later referenced in an assignment.

A DISCONNECT request may be used to cause a CONNECTed PORT to refer once again to the datalanguage input or output port.

```
DISCONNECT <pathname> ;
```

Two CONNECT requests may be issued for the same PORT without an intervening DISCONNECT.

Additional synchronization messages are generated at the time a CONNECTed PORT is used in an assignment statement. These messages are

```
.1230 OCPBO: OPENING INPUT PORT
;1239 OCPBO: INPUT PORT OPENED
.1250 OCPBC: CLOSING INPUT SOCKET
.1240 OCPOO: OPENING OUTPUT PORT
```

```
;I249 OCPOO: OUTPUT PORT OPENED
.I260 OCPOC: CLOSING OUTPUT SOCKET
```

When a CONNECTed PORT is used on the right-hand side of an assignment (that is, in READ mode), the .I230 message is sent over the datalanguage output port. This signals the user program that the datacomputer is attempting to open a network connection to the host and socket specified by the CONNECT request for the PORT. The user program should thus open its end of the connection itself (if it is a connection to a different socket on the user program's own host) or ensure that the third host opens its end of the connection at this time (if it is a connection to another host on the network).

The ;I239 message indicates that indeed the network connection was opened correctly. After this message is received, data can be transmitted, terminated by closing the network connection. Once the connection is closed, the datacomputer sends .I250 over the datalanguage output port, signaling the user program that use of the secondary network connection is complete. The .I250 may precede or follow the closing of the connection on the user's side.

The messages for output PORTs work similarly, with .I240 signaling that the output network connection is being opened, ;I249 that the connection is opened, and .I260 that output is complete and the connection is being closed.

If there are errors in the data, other messages will be sent before the .I250 or .I260 message. This would be the case, for example, if the data does not match the description.

A user program can interrupt the datacomputer's transmission of data; see Appendix D for details.

The form CONNECT <pathname> TO <local-file-designator>; may be useful to those with large amounts of data to send to the datacomputer. In some cases, the shipment of magnetic tapes by air-freight produces higher bit rates than sending the data over the network; the magnetic tape may then be addressed from datalanguage as a local file. Contact CCA for information on this procedure.

## Error Messages

Datacomputer error messages will in general be seen by a human user, although they have header characters which make them potentially processable by a smart user program. Error messages fall into several categories, distinguished by their first character.

| First Character | Meaning |
|---|---|
| ? | indicates a datacomputer or system bug. A user program should rarely see one of these. |

Examples:
```
  ?U000 TRDN: NODE CHAIN SNAFU
  ?U000 DKWR: DISK I/O WRITE ERROR
```

| | |
|---|---|
| - | indicates a user error -- typically bad datalanguage, data, or i/o handling. A debugged user program should rarely see one of these. |

Examples:
```
  -U000 LPNM: FORARG NOT DIRECT LIST MEMBER
  -1246 OCSOP: CAN'T OPEN OUTPUT PORT (BAD CONNECT ARGS?)
```

| | |
|---|---|
| + | indicates a circumstantial error, such as a file's being busy, or an error which is due to current datacomputer limitations. |

```
Examples:      +U000 OCDOP: CAN'T OPEN FILE (SOMEBODY ELSE UPDATING?
  +L000 DHIN: DESCRIPTOR TOO LARGE
```

After the datacomputer generates one or more error message, it follows a special procedure to resynchronize itself with the user. This procedure involves waiting for a special character, control-L or form feed (ASCII 14 octal), to be transmitted by the user. That is, after the error message the datacomputer sends

```
        .1220 LAEB: LOOKING FOR CONTROL-L
```

This is repeated for each line of input it receives on the datalanguage input port until the user sends a control-L character. Following receipt of a control-L, .1210 will again be sent and datalanguage requests again processed.

More severe action must be taken following certain system or ?-type errors. One of the following synchronization messages may be generated:

```
.J151 FCERRH: RESTARTING THE REQUEST HANDLER
.J140 FCREIN: REINITIALIZING USER JOB
.J910 FCERRH: CRASHING JOB
```

The .J151 message indicates that TEMP PORTs have been deleted; otherwise, the status of the session remains the same (PORTs and FILEs will still be open, etc.). This message will usually be followed by .I220, a request for control-L.

The .J140 message is more serious. The user's job is completely reinitialized, leaving his status the same as when the session was begun. This message will also be followed by .I220.

The .J910 message indicates a condition so severe that the datacomputer does not know how to recover. The user's job is crashed and the datalanguage network connections closed. That is, the session is forcibly ended.

If this happens, and also if the user's network connections to the datacomputer are accidentally broken, the datacomputer will do its best to close his open PORTs and FILEs in an orderly manner. However, if the user was in the process of transmitting data into a FILE, the last few thousand characters of data his program sent may have been lost in transit and not incorporated into the FILE.

Not much in general can be said about handling ? or - errors, except that a human user will have to read and interpret the text of the error message in each case, and (in the case of - errors) correct the datalanguage he is having his program send.

+ errors, on the other hand, could be processed by a user program. The most reasonable thing to do in many cases is to wait five minutes and retry the datalanguage request that caused the error. For example, a FILE which was busy (i.e. in use by someone else) may be free by that time, so the second attempt to use it may be successful.

Messages beginning with +L are an exception to this, in that the appropriate time to wait may be several weeks instead of minutes. Such messages indicate limitations of the current datacomputer system, such as limitations imposed by internal table sizes. A new version of the datacomputer may remove many of these limitations. Realistically, this means that +L messages are like - messages in that a program probably could not handle them.

## Appendix A: Summary of Datalanguage Syntax

The following is the complete BNF (Backus Normal Form) specification of datalanguage syntax for version 0/10 of the datacomputer.

### Requests

```
<request>  ::=  ;
```

#### Directory Requests

```
<request>  ::=  LOGIN <login body> ;
<request>  ::=  CREATE <create body> ;
<request>  ::=  DELETE <delete body> ;
<request>  ::=  OPEN <open body> ;
<request>  ::=  CLOSE <close body> ;
<request>  ::=  CONNECT <connect body> ;
<request>  ::=  DISCONNECT <disconnect body> ;
<request>  ::=  MODE <mode body> ;
<request>  ::=  CREATEP <createp body> ;
<request>  ::=  DELETEP <deletep body> ;
<request>  ::=  LIST <list body> ;
```

#### Data Transfer Requests

```
<request>  ::=  <direct assignment> ;
<request>  ::=  <for loop> ;
```

## Directory

### Pathnames

```
<pathname>  ::=  <complete pathname>
<pathname>  ::=  <simple complete pathname>
<pathname>  ::=  <login pathname>
<pathname>  ::=  <simple login pathname>
<pathname>  ::=  <open node name>

<node name>  ::=  <identifier>
<node name>  ::=  <identifier> ( <password string> )
<password string>  ::=  <string constant>
<simple node name>  ::=  <identifier>

<complete pathname>  ::=  %TOP . <node name>
<complete pathname>  ::=
   <complete pathname> . <node name>

<simple complete pathname>  ::=
   %TOP . <simple node name>
<simple complete pathname>  ::=
   <simple complete pathname> . <simple node name>

<login pathname>  ::=  <node name>
<login pathname>  ::=  <login pathname> . <node name>

<simple login pathname>  ::=  <simple node name>
<simple login pathname>  ::=
   <simple login pathname> . <simple node name>

<open node name>  ::=  <simple node name>

<node pathname>  ::=  <complete pathname>
<node pathname>  ::=  <login pathname>

<open pathname>  ::=  <simple complete pathname>
<open pathname>  ::=  <simple login pathname>
<open pathname>  ::=  <open node name>
```

## Directory

### Requests

```
<login body>  ::=  %TOP
<login body>  ::=  <node pathname>

<create body>  ::=  <simple node name>
<create body>  ::=
   <node pathname> . <simple node name>
<create body>  ::=  <data description>
<create body>  ::=
   <node pathname> . <data description>

<delete body>  ::=  **
<delete body>  ::=  <login pathname>
<delete body>  ::=  <login pathname> . **

<open body>  ::=  <node pathname>
<open body>  ::=  <node pathname> <mode>

<close body>  ::=  %OPEN
<close body>  ::=  <open pathname>

<connect body>  ::=
   <open pathname> <tenex file specification>
<connect body>  ::=
   <open pathname> <network specification>
<tenex file specification>  ::=  <string constant>
<network specification>  ::=  <socket number>
<network specification>  ::=
   <host specification> <socket number>
<socket number>  ::=  <integer constant>
<host specification>  ::=  <integer constant>
<host specification>  ::=  <identifier>
<host specification>  ::=  <string constant>

<disconnect body>  ::=  <open pathname>

<mode body>  ::=  <open pathname> <mode>
<mode>  ::=  READ
<mode>  ::=  WRITE
<mode>  ::=  APPEND
<mode>  ::=  WRITE DEFER
<mode>  ::=  APPEND DEFER
```

```
<createp body>  ::=  <node pathname>
<createp body>  ::=
   <node pathname> <privilege tuple specification>
<privilege tuple specification>  ::=
   <privilege tuple option>
<privilege tuple specification>  ::=
   <privilege tuple specification>
           <privilege tuple option>
<privilege tuple option>  ::=  , U = <user identity>
<privilege tuple option>  ::=  , H = <host identity>
<privilege tuple option>  ::=  , S = <socket identity>
<privilege tuple option>  ::=  , P = <password string>
<privilege tuple option>  ::=
   , G = <grant privilege list>
<privilege tuple option>  ::=
   , D = <deny privilege list>
<privilege tuple option>  ::=
   , N = <privilege tuple index>
<user identity>  ::=  **
<user identity>  ::=  <user node>
<user identity>  ::=  <user node set>
<user identity>  ::=  <user node> . **
<user identity>  ::=  <user node set> . **
<user identity>  ::=
   <user node> . <user node set> . **
<user node>  ::=  <identifier>
<user node>  ::=  <user node> . <identifier>
<user node set>  ::=  *
<user node set>  ::=  <user node set> . *
<host identity>  ::=  ANY
<host identity>  ::=  LOCAL
<host identity>  ::=  <integer constant>
<socket identity>  ::=  ANY
<socket identity>  ::=  <integer constant>
<grant privilege list>  ::=  <grant privilege>
<grant privilege list>  ::=
   <grant privilege list><grant privilege>
<grant privilege>  ::=  C
<grant privilege>  ::=  L
<grant privilege>  ::=  R
<grant privilege>  ::=  W
<grant privilege>  ::=  A
<deny privilege list>  ::=  <deny privilege>
<deny privilege list>  ::=
   <deny privilege list><deny privilege>
<deny privilege>  ::=  R
<deny privilege>  ::=  W
<deny privilege>  ::=  A
<privilege tuple index>  ::=  <integer constant>

<deletep body>  ::=
   <node pathname> <privilege tuple index>
```

```
<list body>  ::=  <list node set>
<list body>  ::=  <list node set> <list option>
<list node set>  ::=  %TOP
<list node set>  ::=  %OPEN
<list node set>  ::=  *
<list node set>  ::=  **
<list node set>  ::=  <open node name>
<list node set>  ::=  <node pathname>
<list node set>  ::=  <node pathname> . *
<list node set>  ::=  <node pathname> . **
<list option>  ::=  %NAME
<list option>  ::=  %DESCRIPTION
<list option>  ::=  %DESC
<list option>  ::=  %SOURCE
<list option>  ::=  %ALLOCATION
<list option>  ::=  %ALLOC
<list option>  ::=  %PRIVILEGE
<list option>  ::=  %PRIV
```

## Data Description

```
<datatype>  ::=  <compound datatype>
<datatype>  ::=  <simple datatype>
<datatype>  ::=  <string>

<compound datatype>  ::=  LIST
<compound datatype>  ::=  <structure>
<structure>  ::=  STRUCTURE
<structure>  ::=  STRUCT

<simple datatype>  ::=  BYTE
<simple datatype>  ::=  <integer>
<integer>  ::=  INTEGER
<integer>  ::=  INT

<string>  ::=  <string type>
<string>  ::=  <string type> <string interpretation>
<string type>  ::=  STRING
<string type>  ::=  STR
<string interpretation>  ::=  ASCII
<string interpretation>  ::=  ASCII8
<string interpretation>  ::=  BYTE
<string interpretation>  ::=  INT
<string interpretation>  ::=  INTEGER
```

```
<data description> ::=
   <simple node name> <function>
           <outermost description>
<function> ::= FILE
<function> ::= PORT
<function> ::= TEMPORARY PORT
<function> ::= TEMP PORT
<outermost description> ::= LIST <description>
<outermost description> ::=
   LIST <compound datatype options> <description>
<outermost description> ::= <string>
<outermost description> ::= <string> <string options>
<outermost description> ::= <description>

<description> ::=
   LIST <dimension> <description>
<description> ::=
   LIST <dimension> <compound datatype options>
           <description>
<description> ::=
   <structure> <descriptions> END
<description> ::=
   <structure> <compound datatype options>
           <descriptions> END
<description> ::= BYTE
<description> ::= BYTE <simple datatype options>
<description> ::= <integer>
<description> ::= <integer> <simple datatype options>
<description> ::= <string> <dimension>
<description> ::=
   <string> <dimension> <string options>
<descriptions> ::= <description>
<descriptions> ::= <descriptions> <description>
```

```
<description option>  ::=  <inversion option>
<description option>  ::=  <byte size option>
<description option>  ::=  <filler option>
<description option>  ::=  <variable length option>
<inversion option>  ::=  , I = D
<inversion option>  ::=  , I = I
<byte size option>  ::=  , B = <integer constant>
<filler option>  ::=  , F = <integer constant>
<filler option>  ::=  , F = '<nonquote character>'
<variable length option>  ::=  , C = 1
<variable length option>  ::=  , P = EOF
<variable length option>  ::=  , P = EOB
<variable length option>  ::=  , P = EOR
<variable length option>  ::=  , D = <integer constant>
<variable length option>  ::=
   , D = '<nonquote character>'

<compound datatype options>  ::=
   <compound datatype option>
<compound datatype options>  ::=
   <compound datatype options>
           <compound datatype option>
<compound datatype option>  ::=  <byte size option>
<compound datatype option>  ::=  <filler option>
<compound datatype option>  ::=
   <variable length option>

<simple datatype options>  ::=
   <simple datatype option>
<simple datatype options>  ::=
   <simple datatype options>
           <simple datatype option>
<simple datatype option>  ::=  <inversion option>
<simple datatype option>  ::=  <byte size option>
<simple datatype option>  ::=  <filler option>

<string options>  ::=  <string option>
<string options>  ::=  <string options> <string option>
<string option>  ::=  <inversion option>
<string option>  ::=  <byte size option>
<string option>  ::=  <filler option>
<string option>  ::=  <variable length option>

<dimension>  ::=  ( <integer constant> )
<dimension>  ::=  ( , <integer constant> )
<dimension>  ::=
   ( <integer constant> , <integer constant> )
```

## Data Transfer

```
<data reference>  ::=  <identifier>
<data reference>  ::=  <data reference> . <identifier>
<constant>  ::=  <string constant>
<constant>  ::=  <integer constant>
<assignment>  ::=  <data reference> = <data reference>
<assignment>  ::=  <data reference> = <constant>

<direct assignment>  ::=  <assignment>
<direct assignment>  ::=  <implicit for loop>
<implicit for loop>  ::=  <assignment> <qualifier>

<for loop>  ::=  FOR <input> <for body> END
<for loop>  ::=  FOR <input> <qualifier> <for body> END
<for loop>  ::=  FOR <output> , <input> <for body> END
<for loop>  ::=
   FOR <output> , <input> <qualifier> <for body> END
<input>  ::=  <data reference>
<output>  ::=  <data reference>
<for body>  ::=  <for loop>
<for body>  ::=  <for loop> ;
<for body>  ::=  <assignment list>
<for body>  ::=  <assignment list> ;
<assignment list>  ::=  <assignment>
<assignment list>  ::=
   <assignment list> ; <assignment>

<qualifier>  ::=  WITH <boolean expression>

<boolean expression>  ::=  <relational expression>
<boolean expression>  ::=  ( <boolean expression> )
<boolean expression>  ::=  NOT <boolean expression>
<boolean expression>  ::=  ANY <boolean expression>
<boolean expression>  ::=
   <boolean expression> AND <boolean expression>
<boolean expression>  ::=
   <boolean expression> OR <boolean expression>

<relational expression>  ::=
   <data reference> <comparison operator>
           <data reference>
<relational expression>  ::=
   <data reference> <comparison operator> <constant>
<comparison operator>  ::=  EQ
<comparison operator>  ::=  NE
<comparison operator>  ::=  GT
<comparison operator>  ::=  GE
<comparison operator>  ::=  LT
<comparison operator>  ::=  LE
```

## Lexical Items

```
<lexical item>  ::=  <identifier>
<lexical item>  ::=  <integer constant>
<lexical item>  ::=  <string constant>
<lexical item>  ::=  <autonomous character>

<identifier>  ::=  <letter>
<identifier>  ::=  %
<identifier>  ::=  <identifier> <letter>
<identifier>  ::=  <identifier> %
<identifier>  ::=  <identifier> <digit>

<integer constant>  ::=  <digit>
<integer constant>  ::=  <integer constant> <digit>

<string constant>  ::=  '<string constant body>'
<string constant body>  ::=  <nonquote character>
<string constant body>  ::=
   <string constant body> <nonquote character>
```

## Character Set

```
<letter>  ::=  A
<letter>  ::=  B
 ..............
<letter>  ::=  Z
<letter>  ::=  a
<letter>  ::=  b
 ..............
<letter>  ::=  z

<digit>  ::=  0
<digit>  ::=  1
 ..............
<digit>  ::=  9

<nonquote character>  ::=  <letter>
<nonquote character>  ::=  %
<nonquote character>  ::=  <digit>
<nonquote character>  ::=  <autonomous character>
<nonquote character>  ::=      (space)
<nonquote character>  ::=      (horizontal tab -- HT)
<nonquote character>  ::=  "'
<nonquote character>  ::=  ""

<separator>  ::=      (space)
<separator>  ::=      (horizontal tab -- HT)
<separator>  ::=  <eol>
<eol>  ::=      (end of line -- octal 37)
<eol>  ::=  <carriage return> <line feed>
<carriage return>  ::=      (carriage return -- CR)
<line feed>  ::=      (line feed -- LF)
```

```
<autonomous character>  ::=  !
<autonomous character>  ::=  #
<autonomous character>  ::=  $
<autonomous character>  ::=  &
<autonomous character>  ::=  (
<autonomous character>  ::=  )
<autonomous character>  ::=  *
<autonomous character>  ::=  +
<autonomous character>  ::=  ,
<autonomous character>  ::=  -
<autonomous character>  ::=  .
<autonomous character>  ::=  /
<autonomous character>  ::=  :
<autonomous character>  ::=  ;
<autonomous character>  ::=  <
<autonomous character>  ::=  =
<autonomous character>  ::=  >
<autonomous character>  ::=  ?
<autonomous character>  ::=  @
<autonomous character>  ::=  "B
<autonomous character>  ::=  "/
<autonomous character>  ::=  "E
<autonomous character>  ::=  "!
<autonomous character>  ::=
<autonomous character>  ::=  π*
<autonomous character>  ::=  "6
<autonomous character>  ::=  |
<autonomous character>  ::=  "9
<autonomous character>  ::=  ¬
```

## Notes

Character codes are 7 bit ASCII.

Separators are always permitted between lexical items, except between grant privileges, between deny privileges, and inside string constants.

Comments may be inserted wherever separators are allowed. Comments begin with '/*' and end with '*/' (e.g., /* THIS IS A COMMENT */).

<carriage return> and <line feed> may only appear together in that order (as an <eol>). Otherwise they are treated as control characters, which are rejected.

## Appendix B: Reserved Words

AND
ANY
ASCII
ASCII8
BYTE
CLOSE
CONNECT
CREATE
CREATEP
DELETE
DELETEP
DISCONNECT
END
EQ
FILE
FOR
GE
GT
INT
INTEGER
LE
LIST
LOGIN
LT
MODE
NE
NOT
OPEN
OR
PORT
STR
STRING
STRUCT
STRUCTURE
WITH
%OPEN
%TOP

## Appendix C: Inversion: Technical Considerations

An inversion is a secondary data structure that the datacomputer can use to improve its efficiency in retrieving data by content from a datalanguage FILE. Specifically, an entry in the inversion is constructed for every STR with the inversion attribute. For each data value which occurs for the STR, the inversion contains pointers to all the records in the FILE for which that STR contains that value.

For example, if

```
  CREATE PEOPLE FILE LIST
        PERSON STRUCT
            NAME STR (15)
            SOCSECNO STR(9),I=D
            SEX STR (1) /* 'M' OR 'F'*/,I=D
            ZIP STR(5),I=D
        END;
```

then the data structure for the inversion on SEX contains pointers to all instances of PERSONs with SEX equal to 'F', and similarly for 'M'. Thus, evaluation of a simple FOR input-spec like

```
        FOR ... , PEOPLE.PERSON WITH SEX EQ 'M'
```

would be quick and simple, and would require only a read of the inversion, not any reading of the FILE PEOPLE itself.

An inversion is not only constructed automatically by the datacomputer when the FILE is loaded with data, but is automatically maintained (updated) whenever information in the FILE is updated.

Unfortunately, even if an inversion for the appropriate STR exists, the datacomputer cannot always use it for the evaluation of input-specs, and must sometimes resort to time-consuming searches of the FILE. In particular, the inversion can be used only when the STR is compared with a constant <u>using the operators EQ and NE</u>. That is,

```
          PEOPLE.PERSON WITH ZIP EQ '02138' OR ZIP EQ
 '02139'
                        OR ZIP EQ '02140' OR ZIP EQ
 '02141'
```

can be evaluated directly from the inversion. However,

```
          PEOPLE.PERSON WITH ZIP GE '02138'
                    AND ZIP LE '02141'
```

while it still can be evaluated, cannot take advantage of the inversion and so would be much less efficient datalanguage.

Furthermore, when the STR is a member of an inner LIST, only the oehrator EQ can be evaluated using the inversion. A sequential search is used for evaluating NE.

Complex Boolean expressions, those involving several comparisons, fall into three classes: those with all comparisons evaluable from the inversion, those containing no comparisons evaluable from the inversion, and those which mix the two kinds of comparisons. The first two classes pose no problem; the datacomputer will use the inversion to evaluate expressions in the first category, and not for expressions in the second category.

For mixed expressions, the datacomputer will use the inversion as much as it can. For the present, this can be stated as follows: if the Boolean expression is of the form

```
          <expr> AND <expr> AND ...
```

(where <expr> is an arbitrary Boolean expression, in parentheses if it contains OR) then the datacomputer will separate the <expr>s into those that can be completely evaluated from the inversion and those that cannot, and will process those that can use the inversion first. The <expr>s that cannot use the inversion are evaluated by an exhaustive search of the set of records selected by the earlier <expr>s.

For an example, take the above FILE, PEOPLE. Suppose a list of all males with ZIP GT '02000' were desired. ZIP is indeed inverted, but since the operator GT is involved, the evaluation of that part of the Boolean expression cannot use the inversion. As a result, in

```
          FOR  ... , PEOPLE.PERSON WITH  ZIP GT '02000'
     AND SEX EQ 'M'
```

the datacomputer will first use the inversion to find the set of all PERSONs with SEX EQ 'M', and only this smaller set of PERSONs would be searched for the desired ZIPs.

A more difficult example: consider the problem of retrieving all the records for events that occurred between 10:05 on the 25th and 15:07 of the 30th from a FILE that is inverted on DAY but not on TIME. A straightforward way to do this is

```
              ... WITH (DAY EQ '25' AND TIME GT '10:05')
                     OR (DAY EQ '26') OR (DAY EQ '27') OR
     ...
                      OR (DAY EQ '30' AND TIME LT '15:07')
```

but this is quite inefficient: the inversion cannot be used at all, for this Boolean expression is mixed and is not set up as a series of terms connected by AND. The best way to express this condition is

```
            ... WITH (DAY EQ '25' OR DAY EQ '26' OR ... OR
     DAY EQ '30')
                    AND (DAY NE '25' OR TIME GT '10:05')
                       AND (DAY NE '30' OR TIME LT '15:07')
```

In this case, only records for the correct six days are retrieved by the first term, so only they need to be searched through for the evaluation of the second and third terms.

Future versions of the datacomputer will automatically optimize mixed Boolean expressions, freeing the user from this task.

The computation of the space requirements for an inversion is best left to the datacomputer's operational staff at CCA, who should be contacted by any user interested in setting up a data file with an inversion.

## Appendix D: Network Interaction with the Datacomputer

The procedure for establishing network connections with the datacomputer is that documented in J. Postel, Official Telnet - Logger Initial Connection Protocol, NIC 7103, 15 June 1971. The following is a simplified, informal description of that procedure.

The datacomputer listens for connections on a well-advertised socket, currently number 103 (octal) at CCA, host number 37 (octal). This is an odd-numbered or send socket. The user program wishing to use the datacomputer will address this socket from a socket on his own host computer -- say from socket number U. U must, of course, be an even number or a receive socket. The user program should read one 32-bit byte of information over this connection and then immediately close it (leaving socket CCA-103 free for other users). This byte of information is a socket number at the datacomputer -- say socket D. D will be an even number.

The last step is the opening of two network connections, the permanent datalanguage connections. They are

```
        from D+1 at CCA to U+2 at the user host
  and   from U+3 at the user host to D at CCA.
```

Note that U+2 is even (since U is) and D+1 is odd -- this is the datalanguage output socket. Also, U+3 is odd, and D is even: the datalanguage input socket. These connections will remain in effect until the end of the datalanguage session.

The byte size of the permanent datalanguage connections is 8 bits. The datacomputer sends, and expects to receive, 7-bit ASCII characters right-justified in 8-bit bytes.

Two special network control signals, INS and INR, may be used to interrupt the datacomputer. INS, for Interrupt the sender, may be sent at any time during the processing of a request and stops data output from the current request. No error message or other acknowledgement will be generated; the output simply stops. INS might be useful to a program which receives output fr[illegible] the datacomputer and displays it to a human operator sitting at a teletype; at the request of the user, the program could send INS to stop an overly-long printout.

INR, for interrupt the receiver, performs all the functions of INS. In addition, compilation or any other processing that is under way when INR is received will be aborted, possibly generating an error message and a request for control-L. INR thus requests a more immediate halt than does INS.

## Appendix E: Implementation Restrictions

A number of datalanguage restrictions specific to Version 0/10 are collected here for ready reference. Note that some of these restrictions have been mentioned in the body of this manual, while others have not.

1. There is a restriction on the containers that can be referenced in the body of a FOR-loop. Consider the following example:

```
   CREATE FF FILE LIST
       PERSON STRUCT
           NAME STR (15)
           ADDRESS STR (20)
           CITY STR (10)
           STATE STR (2)
           ZIP STR (5)
           SOCSECNO STR (10)
           DEPENDENTS LIST (10)
               NAME STR (15)
       END;

   CREATE PP PORT LIST
       PERSON STRUCT
           NAME STR (15)
           SOCSECNO STR (15)
       END;
```

To output all the DEPENDENTS.NAMEs from the file FF, together with the SOCSECNO of the PERSON whose DEPENDENTS they were,

```
   FOR PP.PERSON,FF.PERSON
       NAME=NAME;
       SOCSECNO=SOCSECNO;
   END;
```

This example as written will work in datalanguage 0/10. However, if SOCSECNO occurred after DEPENDENTS in the description of FF.PERSON, the request would fail due to a compiler restriction.

When an inner FOR-loop is processing a LIST which occurs within a STRUCT, references may be made in the body of that FOR to objects which occur before that LIST in the STRUCT, but not after the LIST.

There are certain cases of assignment involving inner LISTs which the compiler in Version 0/10 cannot handle. For example, given two structures of the following format:

```
   L1 FILE LIST
       S1 STRUCT
           A1 STR (8)
           A2 LIST (4)
               B2 STR (6)
       END;
```

and

```
   L2 PORT LIST
       S1 STRUCT
           A1 STR (8)
           A2 LIST (4)
               B2 STR (6)
       END;
```

the following FOR-loop will not work:

```
   FOR L1.S1,L2.S2
       FOR A2.B2,A2.B2
           S1=S1
       END
   END;
```

The A2 lists are in use by the inner FOR-loop (FOR A2.B2,A2.B2) when the assignment S1=S1 is encountered. The datacomputer expands S1=S1 internally into:

```
   A1=A1
   FOR L1.S1.A2.B2,L2.S1.A2.B2
       B2=B2
   END;
```

This constitutes a second use of the A2 lists, which cannot be handled.

2. In Version 0/10 of datalanguage, there is one general restriction on sequences of nested FOR-loops, which can be stated as follows:

Sequences of nested FOR-loops are restricted to be a number (possibly 0) of FOR-loops without output LISTs, followed by an arbitrary number, at least 1, of FOR-loops with output LISTs.

For example,

```
    FOR A                    FOR A                  FOR A
      FOR B,C                  FOR B
 (ASSIGNMENT)
        (ASSIGNMENT)             FOR C,D            END;
      END;                         (ASSIGNMENT)
    END;                         END;
                               END;
                             END;
```

The first two examples are legal, whereas the third is not.

3. A FOR-loop with no output LIST can contain only one datalanguage statement as the FOR-body, not a series of statements. Because of restriction 2, that one statement must be a FOR.
This does not apply to a FOR with an output LIST.

4. The only comparison operators which can be evaluated from an inversion are EQ and NE. All other comparison operators must be evaluated by a linear search through a set of records. If the container being compared is a member of an inner list, only the EQ comparison operator can be evaluated from an inversion.

5. It is impossible to assign members of a LIST without setting up a FOR-loop (either explicitly or implicitly). For example, given the PORT is:

```
   CREATE L1 PORT LIST (5)
      S1 STR (3);
```

The following assignment is illegal:

```
   L1.S1='FOO';
```

because it treats the five members of S1 as if they were a single data item.

6. Two outermost containers with the same name may not be open at the same time. This is true even though the containers may have different pathnames in the directory.

7. If an output PORT is punctuated, all assignments before each punctuation character must be completed before any assignments are made after the punctuation character. That is, the datacomputer cannot back up over punctuation in an output PORT. For example, given an output PORT of the form:

```
   PP PORT LIST
      S1 STRUCT
         A1 STR (3),P=EOR
         A2 STR (3),P=EOR
      END
```

assignments must be made in the same order as the STRs appear in the STRUCT.

```
   A1='FOO';
   A2='BAR';
```

will take effect correctly, but

```
   A2='BAR';
   A1='FOO';
```

will not.

Because of the internal paging of the datacomputer, STRUCTs containing long STRs (i.e. greater than 2560 ASCII characters) have a similar restriction. for example, the LIST

```
FF FILE LIST
    S1 STRUCT
        A1 STR (10000)
        A2 STR (10000)
        A3 STR (10000)
    END
```

may have assignments done only in the same order as they appear in the STRUCT.

## Appendix F: Differences between 0/9 and 0/10

The following is a list of changes which, when performed on 0/9 datalanguage, results in the datalanguage for 0/10. The changes are purely user specifiable (i.e. syntactic) features.

### Additions

Login
- The LOGIN request
- The login context
- The %TOP context

Privileges
- The CREATEP request
- The DELETEP request
- Passwords in pathnames

Simple pathnames (without passwords) for open nodes

Variable length
- Data description options -- P, D, C
- Dimension -- (min,max)

The datatype BYTE

The implicit FOR loop

The boolean operator ANY

Defer mode -- WRITE DEFER, APPEND DEFER

New data description options -- I=I, B, F

New LIST options -- %ALLOCATION, %PRIVILEGE

LIST %TOP

LIST *

CLOSE %OPEN

String interpretations -- ASCII, ASCII8, BYTE

Synonyms -- STRUCTURE, STRING

## Modifications

** replaces %ALL

Elimination of the . between LIST's nodes and option

%NAME is now an explicit LIST option

DELETE <pathname>.** is now explicit

MFG. 1-74

CAUTION — REMOVE PROTECTOR SHEET BEFORE TYPING
"TO BE STORED IN A COOL DRY LOCATION"

| | | | | |
|---|---|---|---|---|
| **MATERIAL INSPECTION AND RECEIVING REPORT** | 1. PROC. INSTRUMENT IDEN(CONTRACT) MDA903-74-C-0225 | (ORDER) NO. | 6. INVOICE NO. DATE | 7. PAGE 1 OF 1; 8. ACCEPTANCE POINT D |
| 2. SHIPMENT NO. CCA00003 | 3. DATE SHIPPED 9/20/74 | 4. B/L TCN | 5. DISCOUNT TERMS | |

| | |
|---|---|
| 9. PRIME CONTRACTOR CODE 6A046<br>Computer Corporation of America<br>575 Technology Square<br>Cambridge, Massachusetts 02139 | 10. ADMINISTERED BY CODE S2202A<br>Mr. J. McDonough<br>Defense Contract Administration<br>Services Region, Boston<br>666 Summer Street<br>Boston, Massachusetts 02210 |
| 11. SHIPPED FROM (If other than 9) CODE FOB:<br>Same as 9. above | 12. PAYMENT WILL BE MADE BY CODE S2202A<br>Disbursing Officer<br>Defense Contract Administration<br>Services Region, Boston<br>666 Summer Street<br>Boston, Massachusetts 02210 |
| 13. SHIPPED TO CODE W73QQB<br>Defense Advanced Research Projects Agy.<br>Architect Building<br>1400 Wilson Boulevard<br>Attn: Dr. Craig Fields<br>Arlington, Virginia 22209 | 14. MARKED FOR CODE<br>Same as 13. |

| 15. ITEM NO. | 16. STOCK/PART NO. DESCRIPTION (Indicate number of shipping containers - type of container - container number.) | 17. QUANTITY SHIP/REC'D * | 18. UNIT | 19. UNIT PRICE | 20. AMOUNT |
|---|---|---|---|---|---|
| 0002 AB | Semi-Annual Technical Report | 2 | | | NSP |

21. PROCUREMENT QUALITY ASSURANCE

A. ORIGIN

☐ PQA ☐ ACCEPTANCE of listed items has been made by me or under my supervision and they conform to contract, except as noted herein or on supporting documents.

DATE SIGNATURE OF AUTH GOVT REP

TYPED NAME AND OFFICE

B. DESTINATION

☐ PQA ☐ ACCEPTANCE of listed items has been made by me or under my supervision and they conform to contract, except as noted herein or on supporting documents.

DATE SIGNATURE OF AUTH GOVT REP

TYPED NAME AND TITLE

22. RECEIVER'S USE

Quantities shown in column 17 were received in apparent good condition except as noted.

DATE RECEIVED SIGNATURE OF AUTH GOVT REP

TYPED NAME AND OFFICE

* If quantity received by the Government is the same as quantity shipped, indicate by ( ✓ ) mark, if different, enter actual quantity received below quantity shipped and encircle.

23. CONTRACTOR USE ONLY