NASA

N91-27025

NASA/SPAN AND DOE/ESNET-DECNET
TRANSITION STRATEGY FOR DECNET OSI/PHASE V

Linda Porter and Phil DeMar*

Space Science Laboratory
National Aeronautics And Space Administration
George C. Marshall Space Flight Center
Huntsville, Alabama

SSL

MSFC - Cover 32 (Rev. November 1985)

NASA/SPAN AND DOE/ESNET-DECNET TRANSITION STRATEGY FOR DECNET OSI/PHASE V

Linda Porter, NASA/Marshall Space Flight Center
Phil DeMar, DOE/Fermi National Accelerator Laboratories

ABSTRACT

This paper examines the technical issues involved with the transition of very large DECnet networks from DECnet Phase IV protocols to DECnet OSI/Phase V protocols. The networks involved are the National Aeronautics and Space Administration's Science Internet (NSI-DECnet) and the Department Of Energy's (DOE's) Energy Sciences network (ESnet-DECnet). These networks, along with the many universities and research institutions connected to them, combine to form a single DECnet network containing more than 20,000 nodes and crossing numerous organizational boundaries. The transition planning for this network must deal with both the scale of the network and its administrative complexity. This necessitates creation of a transition strategy that is flexible enough to allow different parts of the network to upgrade to Phase V at different times, yet is sufficiently coordinated so that network functions are not disrupted.

Discussion of transition planning, including decisions about Phase V naming, addressing, and routing are presented. Also discussed are transition issues related to the use of non-DEC routers in the network.

INTRODUCTION

The DECnet Internet is a very large DECnet-based network reaching government, university and research sites throughout the world which are involved in scientific research. The network has grown from numerous small, disconnected DECnet networks of 10 years ago to a conglomerate network which crosses numerous international and organizational boundaries. The DECnet Internet, therefore, is not an "engineered" network, rather, it is the result of the growth and interconnection between a number of smaller, previously independent DECnet networks.

The four largest participants in the DECnet Internet are the NSI-DECnet (formerly SPAN), the ESnet-DECnet, the European Space Agency's Space Physics Analysis Network (E-SPAN) and the consortium of European High-Energy Physics Research Institutions (E-HEPnet). Other participants include scientific DECnet networks in Japan, Canada, South America, and Australia. Administratively separate, these DECnet networks share a common address space and lie within a single routing domain. The result is a single huge DECnet network of thousands of nodes, complicated architecture and many network managers.

In the U.S., the NSI-DECnet and the ESnet-DECnet comprise the bulk of the DECnet Internet systems:

- The NSI-DECnet is a NASA-funded network supporting space plasma physics and astrophysics as well as related space science research programs. NSI-DECnet reaches more than 80 sites, including most of the NASA field centers and universities that are involved in NASA research programs. The network also has connections to other DECnet networks throughout the world that engage in space science research and programs (Figure 1).

- The ESnet-DECnet is a DOE-funded network supporting energy research programs such as high energy physics, nuclear physics, and fusion research. It connects together over 60 sites in the United States, including the major national laboratories, as well as universities involved in energy research programs. The ESnet-DECnet, like the NSI-DECnet, supports numerous connections to other DECnet networks around the world involved in energy research (Figure 2).

The network management teams for the four major participant networks coordinate operations through the "HEP-SPAN DECnet Coordinating Group", or HSDCG, to ensure the network functions properly. The HSDCG is involved in coordinating technical issues such as address usage and circuit cost assignments (routing), as well as administrative issues such as security incident handling and network information distribution. The primary task now facing the group is planning for the transition of DECnet Phase IV protocols in use on the network today to DECnet OSI/Phase V.

Complicating the planning for implementing Phase V on the DECnet Internet are the numerous interconnections (dashed lines) between the networks (Figure 3). These interconections were originally installed to serve specific program or research requirements rather than improve overall network performance. There are no less than 17 interconnections between NSI-DECnet and ESnet-DECnet in the U.S. Although these links provide redundancy, they also add many routers to the network, making the routing topology very complicated and the transition planning more difficult. As we'll see later on, routers are key elements in the transition.

This paper deals with the major issues involved in planning for the transition to DECnet OSI/Phase V, primarily from the perspective of the NSI-DECnet and ESnet-DECnet networks. First we examine the motivations behind the requirement to use Phase V protocols. Next we present constraints on the transition planning, including a discussion on maintaining Phase IV connectivity and implementing OSI protocols for the anticipated future network environment. We then outline the general transition strategy for the DECnet Internet. Finally, we present a technical discussion of OSI/Phase V addressing, naming and routing issues.

NSI / Space Physics Analysis Network

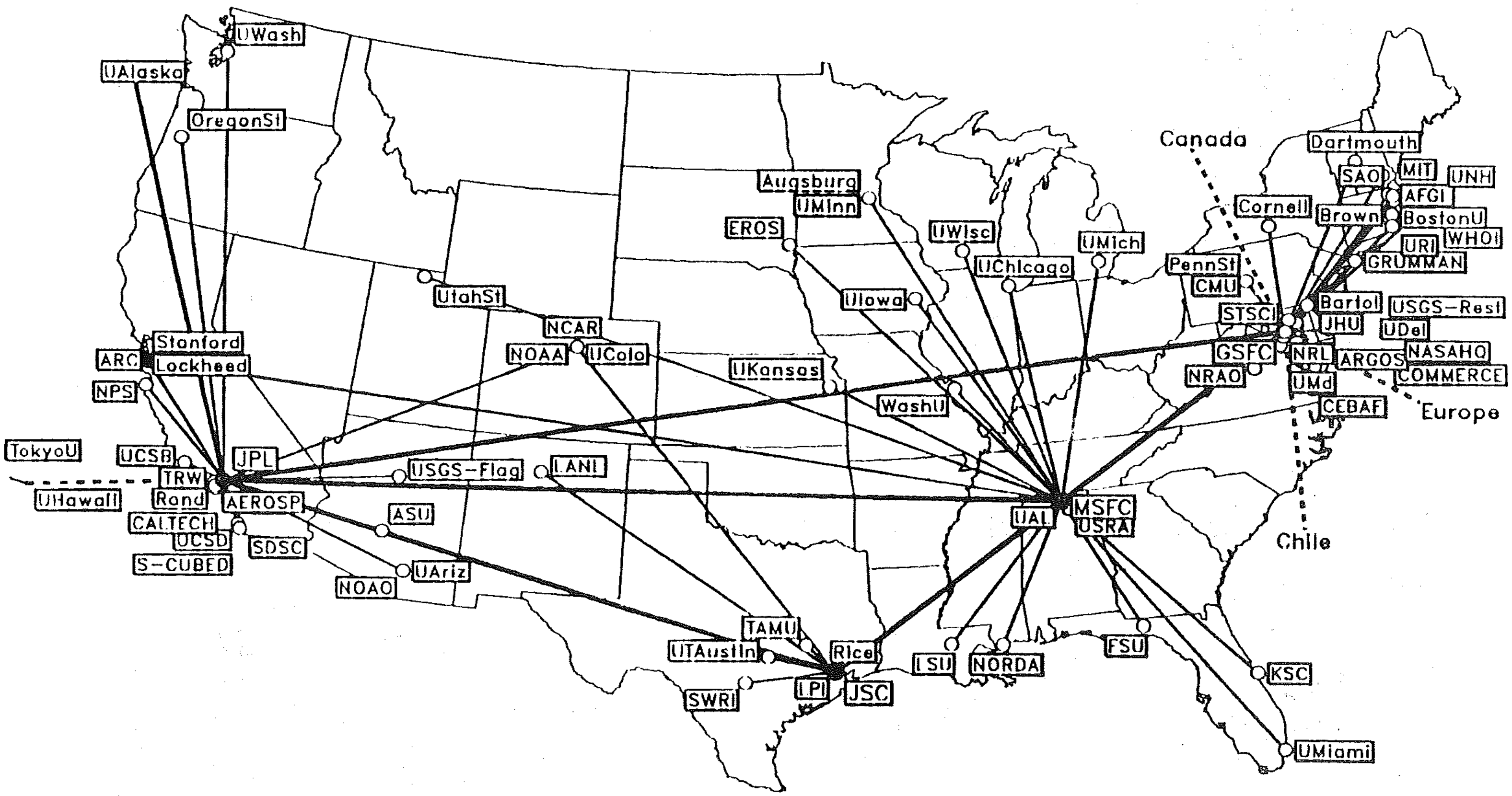


FIGURE 1.

Prepared for: NASA Science Internet Project Office
by: Sterling Software, NAS2-11555
March 1, 1990

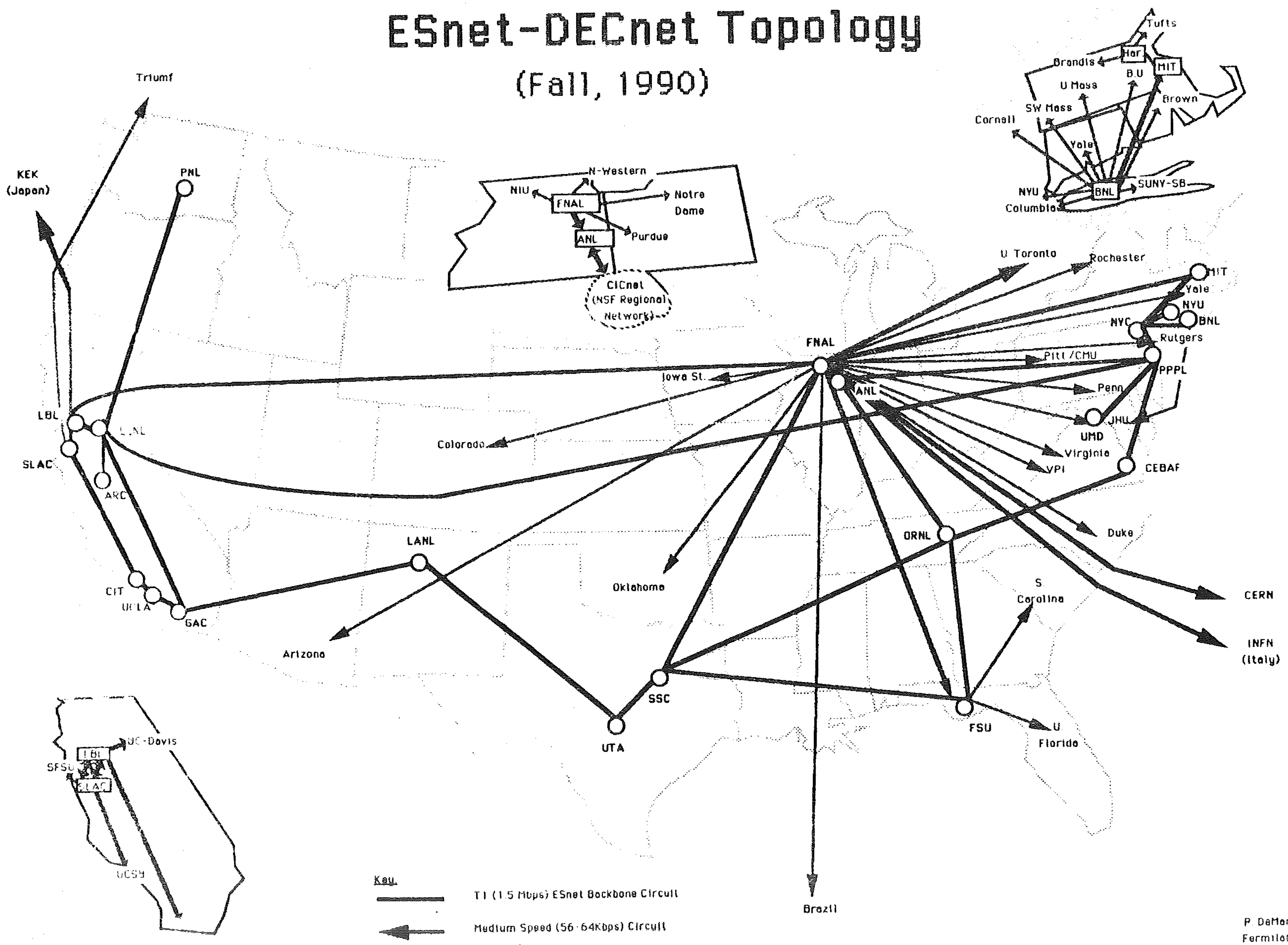

ESnet-DECnet Topology
(Fall, 1990)
Triumf
KEK (Japan)
PNL
LBL
LLNL
SLAC
ARC
CIT
UCLA
GAC
LANL
Arizona
UTA
SSC
Colorado
Iowa St.
FNAL
ANL
Oklahoma
Brazil
ORNL
FSU
U Florida
S Carolina
CERN
INFN (Italy)
Duke
VPI
Virginia
UMD
JHU
CEBAF
Penn
Pitt/CMU
PPPL
Rutgers
NYC
NYU
BNL
Yale
MIT
Rochester
U Toronto
N-Western
NIU
FNAL
Notre Dame
Purdue
ANL
CICnet (NSF Regional Network)
Tufts
Brandis
Har
MIT
B.U
U Mass
Brown
SW Mass
Cornell
Yale
NYU
Columbia
BNL
SUNY-SB
UC-Davis
LBL
SFSU
SLAC
UCSD
Key
T1 (1.5 Mbps) ESnet Backbone Circuit
Medium Speed (56-64Kbps) Circuit
Low Speed (19.2 Kbps or less) Circuit
P. DeMar
Fermilab
10/90

FIGURE 2.

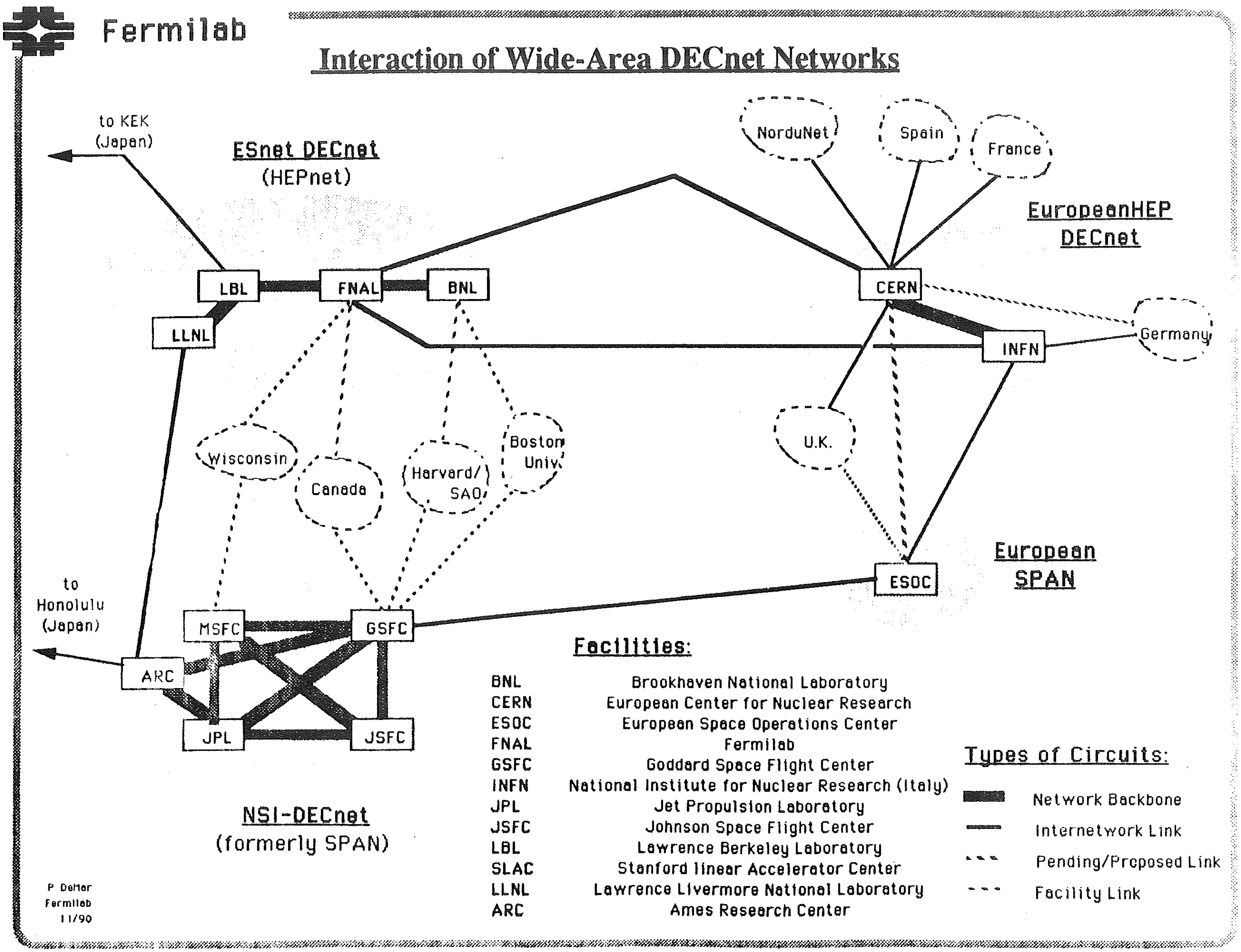

Fermilab
Interaction of Wide-Area DECnet Networks
to KEK
(Japan)
ESnet DECnet
(HEPnet)
NorduNet
Spain
France
EuropeanHEP
DECnet
LBL
FNAL
BNL
CERN
LLNL
Germany
INFN
Wisconsin
Canada
Harvard/
SAO
Boston
Univ.
U.K.
European
SPAN
ESOC
to
Honolulu
(Japan)
MSFC
GSFC
ARC
JPL
JSFC
NSI-DECnet
(formerly SPAN)
Facilities:
BNL Brookhaven National Laboratory
CERN European Center for Nuclear Research
ESOC European Space Operations Center
FNAL Fermilab
GSFC Goddard Space Flight Center
INFN National Institute for Nuclear Research (Italy)
JPL Jet Propulsion Laboratory
JSFC Johnson Space Flight Center
LBL Lawrence Berkeley Laboratory
SLAC Stanford linear Accelerator Center
LLNL Lawrence Livermore National Laboratory
ARC Ames Research Center
Types of Circuits:
Network Backbone
Internetwork Link
Pending/Proposed Link
Facility Link
P DeMar
Fermilab
11/90

FIGURE 3.

WHY DO WE NEED TO UPGRADE TO PHASE V?

The driving force for an early transition to DECnet Phase V is the fact that the DECnet Internet has reached the practical addressing and routing limits of DECnet Phase IV. Also, DECnet Phase V promises integrated support for OSI, which is expected to be the protocol of choice for the future. Finally, the U.S. Government, through its Government OSI Profile (GOSIP) procurement policies will require its networked systems to support the OSI protocols.

PHASE IV ADDRESSING LIMITATIONS

DECnet Phase IV allows only 2 bytes for a node address, which is further divided into 63 areas. The DECnet Internet, which consists of numerous networks and hundreds of sites in many countries, cannot meet its addressing requirements with only 63 DECnet areas. For example, it is very difficult to inform sites in different countries, used to their own autonomy, that they need to share the same DECnet area and coordinate their address assignment policies. In addition, the cost of routing packets over the public switched facilities for European sites is staggering when sites share a single DECnet area and face charges for routine exchange of voluminous intra-area routing information.

Various area filtering techniques have been utilized to deal with the limited address space. These area filtering techniques have created "hidden" areas. Hidden areas are defined as those areas which are intentionally invisible to most of the network. As a consequence, certain area numbers can be duplicated without impact on normal network operations. These hidden areas, however, make network management difficult, and can break the network if the filtering mechanism is accidently removed.

DECnet OSI/Phase V provides 20 bytes of address space, obviously solving the limitations of Phase IV addressing. Just how big is 20 bytes? Well, it's probably enough to assign every toaster (5 billion per planet) on every planet in the universe (about 10E+22) with about 20 quadrillion addresses (a 2 followed by 16 zeros). Although not quite infinity, 20 bytes will probably cover addressing requirements until we retire.

PHASE IV ROUTING PROBLEMS

While Phase IV address space is bounded, Phase IV routing is boundless. This means the entire network is contained within a single routing domain, creating a number of problems:

- The network is very vulnerable to inadvertent connections that bring duplicate area numbers into the network which, unlike hidden areas, are very visible. Visible duplicate area numbers cause network partitioning. In a partitioned network, parts of the network cannot exchange messages with other parts of the network.

- o With a heavily interconnected topology using a single routing protocol, derivation of appropriate DECnet circuit costs for achieving proper traffic flow becomes very complex and very difficult.

- o The numerous routing loops in the network often cause unexpected and inappropriate routing during periods of circuit instability. This causes poor performance, or in some cases, prevents packets from reaching their destination. Routing loops are a consequence of the failure of the Phase IV distance vector routing algorithm when used in a large network with a complex topology, like the DECnet Internet.

DECnet OSI/Phase V provides definable routing domain boundaries and the ability to control what routing information is propagated into or out of any particular network. With such control, inadvertent connections are harder to make, and ease problems of duplicated areas.

Also, the Phase V link state routing algorithm is much more robust and scalable than the Phase IV distance vector algorithm, and it eliminates the Phase IV routing loop problems.

THE TREND TOWARD OPEN NETWORKING PROTOCOLS AND U.S. GOSIP

It is generally accepted that most institutions will use OSI protocols eventually. The International Standards Organization (ISO) is driving the development of OSI protocols for the purpose of providing worldwide computer interoperability.

DECnet OSI/Phase V implements OSI protocols while preserving interoperability with DECnet Phase IV systems. No other protocol can provide for a relatively transparent transition from DECnet Phase IV to OSI (or other) protocols.

Also, the U.S. government mandates specification of OSI for networked systems in purchases. These practices are defined in the Government Open Systems Interconnect Profile (GOSIP) procurement specification (Federal Information Processing Standard 146). GOSIP also describes the OSI protocols to be used, and their formats. The intent is to eventually make all networked Government systems use OSI, resulting in greater interoperability and hence less reliance on any particular computer or network vendor. A significant portion of the network, therefore, will be required to support OSI.

CONSTRAINTS ON PHASE V TRANSITION PLANNING

There are several constraints affecting the development of the transition plan:

- o Backwards compatibility with the existing Phase IV production network must be maintained throughout the transition.

  The transition to Phase V will take an extended period of time, probably requiring several years. During the transition period, Phase IV systems throughout the network must maintain full connectivity with other Phase IV systems and also Phase V systems. In addition, the area filter mechanisms presently used in the Phase IV network must remain until they are either no longer necessary or they can be removed without disrupting the network.

- o Technical constraints on the use of OSI addressing throughout the transition must be understood.

  Because backwards compatibility with Phase IV systems must be maintained, networks are constrained to use Phase IV compatible addresses for Phase V systems during the transition period. Well, it's not surprising the number of Phase IV compatible addresses is identically equal to the number of Phase IV addresses - we still only get to use 2 bytes! The address management and assignment practices presently enforced in the Phase IV network will necessarily remain for assignment of Phase IV compatible addresses during the transition process. However, some systems will be identified as not requiring communication to Phase IV systems during the transition. These may implement their facility-assigned OSI addresses, but not a Phase IV compatible address. After the transition, sole use of the facility-assigned OSI address for all systems will be encouraged.

- o Technical constraints on the use of OSI routing throughout the transition must be understood.

  Phase V allows only one routing algorithm (Phase V or Phase IV) within a specific DECnet area. This means that *all* routers within an area must be able to support Phase V before that particular area can be upgraded. Host-based (VMS) routers present another problem. They will never be able to support the Phase V Level-2 (area) routing, and will probably be somewhat delayed in supporting Phase V Level-1 (intra-area) routing. Note that VMS routers used only for cluster aliasing are not affected. However, facilities using VMS routers for other than cluster aliasing are likely to be severely constrained in efforts to upgrade to Phase V. These sites will be encouraged to move from host-based routers to dedicated routers.

- o The variety of hardware and software in use affects the timing of implementation.

  Allowances for the variety of routers and systems in use in the DECnet Internet must be made in the transition plans. While some parts of the network contain only DEC hardware and software, other parts depend on third-party implementations of DECnet. The planning and timescale for the transition of the latter will almost certainly be different than the former.

- The transition must be implemented in a manner consistent with the long term objective of being part of a global OSI network.

  The new protocols implemented must conform to existing OSI recommendations and specifications. For government sites, GOSIP address formats as well as agency GOSIP transition plans need to be followed. The namespace will be structured to follow the OSI X.500 recommendations, and planned with the idea of becoming part of a global X.500 directory service when that becomes available.

  Routing under OSI must be planned and eventual implementation of "routing domains" consistent with local facility plans must be permitted.

- The organizational complexity of the existing global internet must be considered.

  The DECnet Internet crosses national boundaries as well as agency and facility jurisdictions. Therefore, the transition plan must be flexible enough to meet the differing needs and perspectives of individual facilities and agencies. A top-down approach using a "one strategy fits all" philosophy is very likely to fail miserably.

PHASE V TRANSITION GENERAL STRATEGY

Considering the goals and constraints of the transition, the general strategy for the transition of the DECnet Internet to OSI/Phase V will be based on the following:

- Network backbones are expected to be upgraded to Phase V at the earliest possible time. The underlying philosophy will be "backbone sites first, tail sites last". This provides two things: 1) a central framework around which to base the transition, and 2) upgrade of the major resources on the network at an early time in the transition (since they tend to be located at backbone sites).

- Detailed transition plans for individual networks will generally be based on an area-by-area upgrade - an incremental strategy. Phase IV areas within the DECnet Internet that are ready to upgrade will be identified. These areas will then coordinate a changeover to the use of Phase V protocols all at once. This is not quite as impossible as it sounds, because the primary issue in this changeover is upgrading the *routing* nodes in an area. End systems may run either Phase IV or Phase V software in either a Phase IV or Phase V area. End systems can be upgraded gradually throughout the transition process.

  Two approaches are possible with an area-by-area transition. The first approach identifes the sites within an area ready to upgrade to Phase V. Sites sharing that area which are unprepared or unable to go to Phase V will be assigned new Phase IV addresses and moved, allowing the remaining the sites to proceed with Phase V implementation.

The second approach again starts with identifying sites within an area ready to upgrade to Phase V. This time, though, those sites ready to upgrade will first adopt new Phase IV addresses, thus decoupling them from sites not ready to upgrade. Then the sites with the new addresses will coordinate a changeover to Phase V routing protocols all at once. In some cases, the adoption of new Phase IV compatible address and the changeover to Phase V routing protocols will happen simultaneously.

It is likely that certain areas will remain permanent Phase IV areas to support those systems which will never run OSI protocols.

This incremental strategy provides a means of accelerating the transition process for those portions of the network ready to upgrade. It also provides justification (and motivation) for other sites to hasten their own OSI/Phase V implementations.

o Phase IV backwards compatibility will be preserved by adoption of a common high-order address, or "Phase IV prefix" for all the networks within the DECnet Internet. The common Phase IV prefix will be used to create a virtual routing domain for the Phase IV nodes within the network, preserving the Phase IV address structure. Phase V systems will be multihomed (have multiple addresses) when necessary. On a multihomed system, one address will be the Phase IV compatible address (common Phase IV prefix + existing Phase IV DECnet address). The other address will be the facility assigned OSI address. Addressing is further discussed in the next section.

o There will be a single namespace created to support the Phase V network. Namespace name and structure will be common, and implementation will adhere to guidelines. Directory replication and access, as well as clearinghouse location, will be tightly controlled down to the facility level. The namespace implementation will precede Phase V implementation, and sites will be allowed (encouraged) to utilize the namespace for existing Phase IV applications. Namespace issues are discussed in greater detail in the next section.

o Initially, the number of routing domains in the changing network will be minimized. As the transition progresses, implementation of routing domains will increase. However, there are technical reasons which prevent initial widespread use of routing domains. These reasons are presented in the next section.

o There will be a finite amount of time for completion of the transition across the entire network. After that time ends, the network will be declared a Phase V network, and use of extended address space will be encouraged. Phase IV areas (and Phase IV end systems within Phase V areas) may remain after this time, but direct access to wide-area network resources no longer will be guaranteed. "Poor man's routing" may be required to provide access for those systems.

o ESnet-DECnet, NSI-DECnet, E-HEPnet, E-SPAN, and other network management teams controlling specific parts of the DECnet Internet will each refine its own transition plan, using the transition strategy it deems appropriate for its own network environment. The time scale for each of these individual transition plans will be independent of the others. However, transition strategies and implementation plans will be closely coordinated with other member networks.

TECHNICAL ISSUES FOR DECNET INTERNET TRANSITION PLANNING

The groundwork for understanding the existing network environment, the need for a transition, and the general strategy for the transition has been discussed. The following sections tackle addressing, naming, and routing issues in greater technical detail.

ADDRESSING

The OSI address format to be used by all U.S. Government Institutions is defined by GOSIP. The proper name for this address format is the "Network Service Access Point", or NSAP. The NSAP is 20 bytes long and is shown in Figure 4.

```
<---IDP---> <----------------------DSP------------------------------>
            <------------HO-DSP---------> <-LA-> <-----ID----> <-SEL->
 ---------------------------------------------------------------------
| AFI | IDI | DFI |   AA   | RESV | SNID | AREA | END SYSTEM | NSEL |
 ---------------------------------------------------------------------
   1     2     1      3       2      2      2         6          1    bytes

  47   0005   80   qqqqqq   0000   nnnn   mmmm   abcdefghijkl   yy
                   003400  (NASA)
                   000400  (DOE)
```

IDP : Initial Domain Part
DSP : Domain Specific Part
HO-DSP : High Order Domain Specific Part
LA : Local Area
ID : end system IDentification
SEL : transport SELector byte

AFI : Authority and Format Identifier
IDI : Initial Domain Identifier
DFI : Data Format Identifier
AA : Administration Authority
SNID : Sub-Network ID

FIGURE 4. - THE GOSIP NSAP

GOSIP defines values for the IDP and the DFI. NASA and DOE have applied to the National Institute of Standards and Technology for the value of the "AA" field, and it has been assigned: NASA will use "003400" and DOE will use "000400", as shown in Figure 4. The remainder of the address will be assigned according to internal NASA and DOE recommendations.

AN ADDRESSING PROBLEM

An address field of great interest is defined by the combination of the IDP and HO-DSP fields, or the "high-order address". To explain this, a small digression is needed.

> PHASE V TRANSITION RULE: Phase IV systems can communicate only with systems having the same high-order address as the Phase V routers to which they are connected.

That is, the portion of the address to the left of the Local Area (LA) field must be identical on all Phase V systems if Phase IV end systems are to communicate during the transition. The reason for this is simple: Phase IV systems have no knowledge or ability to generate any address but a Phase IV style address containing an area between 1 and 63 and a node address between 1 and 1023. However, in a Phase V network, Phase IV end-systems actually are assigned a high-order address: it is the high-order address of the Phase V router to which the Phase IV system is connected. But because a Phase IV system itself has no knowledge of its high-order address, it can't generate a different one. Therefore, a Phase IV system can talk only to those systems that are connected to a router with the same high-order address as the Phase V router that is connected to the Phase IV system.

Therefore, the statement of the problem is:

> If all institutions adopt the OSI address format with arbitrary high-order addresses, how can Phase IV system connectivity be maintained?

THE ANSWER

OSI specifies support for multiple addresses for a single system. A system with multiple addresses is said to be "multihomed". If one of the addresses on a Phase V multihomed system contains a prefix common to all other Phase V nodes, then Phase IV connectivity can be preserved. The form of this address is described in Figure 5.

```
<-IDP-> <--------HO-DSP---------------> <----------THE REST--------->
 ____________________________________________________________________
| zz  |            pppp               | AREA | END SYSTEM | NSEL |
 --------------------------------------------------------------------
<----------Phase IV prefix----------->    2        6           1
```

FIGURE 5. - PHASE V/PHASE IV COMPATIBLE ADDRESS FORMAT

This common address can be up to 20 bytes long, and conforms to OSI Standards. (Note that the "AREA:END SYSTEM" must translate to a Phase IV compatible address, i.e. area between 1 and 63, node address between 1 and 1023.)

Therefore, one address on Phase V systems can be GOSIP (or ANSI or other standard). The other address will be the address linking the Phase IV DECnet Internet. For example, a node using Phase IV compatible address 7.39 can have two completely independent addresses as follows:

1. GOSIP COMPLIANT ADDRESS:

IDP	HO-DSP	LA	ID	SEL
47 0005	80 003400 0000 1100	2366	08002b123456	yy

2. PHASE IV/V COMPATIBILITY ADDRESS:

<--------Phase IV Prefix------>

IDP	HO-DSP	LA	ID	SEL
99	4242	0007	aa000400271C	xx

(The "99 4242" is a hypothetical example of a unique Phase IV prefix used in the DECnet Internet for the purpose of transition.)

Therefore, multihomed Phase V systems satisfy requirements for use of OSI while preserving Phase IV compatibility during the transition period.

ADDRESSING AUTHORITY

For the purpose of implementing a Phase IV to OSI/Phase V transition, the existing methods for obtaining *Phase IV* addresses will be unchanged throughout the Phase V transition period. Phase IV addresses will be used with the unique Phase IV prefix to ensure Phase IV/V transparency during the transition. As described above, however, the Phase IV address is unrelated to the value of the facility-assigned OSI address. Sites can receive OSI addresses from their OSI Address Authority at any time (of course).

NAMING ISSUES

The directory and naming service that will be used during the transition is DEC's "Digital Naming Service", or DECdns. DECdns provides, among other things, address-to-name and name-to-address translation services as well as user application and other general naming services. DECdns provides a robust method to keep names and addresses up to date, and a method for replicating portions of the namespace for redundancy. DECdns is expected to interoperate with the OSI X.500 directory service when that becomes available. Network support for DECdns during the transition to DECnet OSI/Phase V is a requirement.

The following definitions are important to understanding naming issues.

Logical namespace - the global structure defining how systems are named.

Physical namespace - the implementation in a working network of the Logical Namespace.

Logical namespace issues are separate from physical namespace issues, and are treated separately.

LOGICAL NAMESPACE ISSUES

The Logical namespace to be used for the DECnet Internet will adhere to OSI X.500 recommendations as closely as possible. It will also be kept as shallow as possible. The general structure of an X.500 name is:

.COUNTRY.ORG.OS...

where ORG is the organization "owning" this specific portion of the namespace, and OS is an "organizational specific" identifier assigned by the owning organization.

NASA's current recommendation for the naming of NASA field centers is the following:

.US.NASA.center.name
e.g.
.US.NASA.MSFC.SSL

DOE's recommendation, and the one now being used in the OSI transition guidelines for that agency is the following:

.US.facility.name	or	.US.DOE.facility.name
e.g.	(for small DOE sites)	e.g.
.US.FNAL.FNMFE		.US.DOE.CHI.name

We can draw three observations from these recommendations:

1) This is backwards to the TCP/IP Internet standard - we don't love it, but if the names are to adhere to X.500 recommendations it is unavoidable that DECnet Phase V system names will be reversed with respect to TCP/IP Internet names.

2) There is no upper-level domain as in the Internet standard, i.e. no "EDU" or "COM" field. The feeling is these fields do not convey useful meaning, and are contrary to the X.500 recommendations.

3) DEC recommends against putting the country symbol in the DECdns namespace for a network. This is because most sites will be joining a larger network - and hence namespace - in the future, where the upper level directories are already provided. This is not suitable for the already international DECnet Internet, where the country code must be present to distinguish international organizations.

DOE and NASA are not naming authorities for X.500 (nobody is, yet!). However, they will recognize and register Internet Facility level domain names, such as "FNAL", "UCSD", and "MSFC" in the namespace for sites currently served by the DECnet Internet.

The intent is to join the DECnet Internet namespace with the global X.500 directory services when available. This will be done by removing the appropriate top level directories in the DECdns namespace and pointing the remainder at the X.500 root. At that time, one presumes, a global naming authority and registration board will exist, and facilities will register with that organization.

PHYSICAL NAMESPACE ISSUES

Institutions such as major DOE sites and NASA field centers will emplace name servers. An invitation to join the logical namespace structure provided by these name servers will be extended to associates. DEC (and we) recommend that there be at least two name servers per local area network.

Each facility joining the namespace will be responsible for maintaining the master copy of its own top level (facility) directory at its local site, just as is presently the case for Internet (TCP/IP) domain name servers. However, read-only copies of facility level directories will likely be located elsewhere in the network as well.

More work needs to be done in deciding guidelines for replication and access of the physical namespace across the DECnet Internet. (Replication assures reachability in case of a network link or server failure.)

ROUTING ISSUES

INTER-DOMAIN VS. INTRA-DOMAIN ROUTING

There is a lot of confusion about inter-domain and intra-domain routing. Many confuse dynamic and static routing issues, and others believe routing hierarchy is many levels deep (it's only two), and routing domains depend on specific fields of the NSAP (they don't). So, sit back, clear your mind, and let's start from scratch.

INTRA-DOMAIN ROUTING

DECnet OSI/Phase V uses a protocol named "IS-IS Routing Exchange Protocol" for intra-domain routing. (IS = Intermediate System, i.e. a router). This protocol is currently at draft international standard stage and will be a full OSI protocol probably within a few months. The IS-IS protocol uses a more robust and scalable routing algorithm than Phase IV called "Link-State Routing". However, the following analogy with DECnet Phase IV will be used to illustrate an important concept.

Like DECnet Phase IV, IS-IS routing has two *and only two* routing levels: Level 1 and Level 2. There is no deeper hierarchical routing specified by this standard. An IS-IS Level 1 router keeps information on every end system in its area, like Phase IV DECnet. An IS-IS Level 2 router keeps information on every other area in the network, again, like Phase IV.

Okay, so what is an OSI area? This is where the NSAP plays a role. The IS-IS standard routes area (Level-2) traffic based on the value of the IDP + HO-DSP + LA fields. Therefore, these fields define the OSI area, as shown in Figure 6.

```
 ___________________________________________________________________________
| IDP    |                          DSP                                     |
|________|__________________________________________________________________|
         |     HO-DSP         |      LA      |         ID          |  SEL   |
         |____________________|______________|_____________________|________|

<----------LEVEL 2 ROUTING------------>
                                        <--LEVEL 1 ROUTING-->
```

Figure 6. - IS-IS ROUTING AND THE NSAP

Now, the amount of space allowed for areas is huge - up to 13 bytes! Instead of being constrained by only 63 areas, an OSI network could wallow in 2.0E31 areas. One can immediately see both the advantages and problems associated with the possibility of tremendous numbers of OSI areas.

INTER-DOMAIN ROUTING

To prevent problems associated with zillions of areas (and for other reasons), network management can define "Routing Domains."

ROUTING DOMAIN: A routing domain is a collection of systems that are told they are running the same routing protocol.

A routing domain can be defined which allows all systems within it to keep their routing information confined, or better - everybody else's routing information out. Defining a routing domain can isolate a group of areas from exchanging routing information with the rest of the world while allowing well-defined interconnection points so that communications between routing domains is still possible. Defining a routing domain, then, can solve two problems. One, it can reduce the number of areas in a network, and two, it can protect a network from routing problems in a neighboring network.

It's important to realize that the mechanism used to define a routing domain is not particularly related to any specific field in the NSAP prefix (the portion of the address above the LA field, Figure 4). There is no special field in the NSAP such that when bits in it are changed, the routing domain is changed as well. A routing domain boundary can be set between any two sites whose NSAP prefixes are different. Conversely, a routing domain can contain multiple NSAP prefixes.

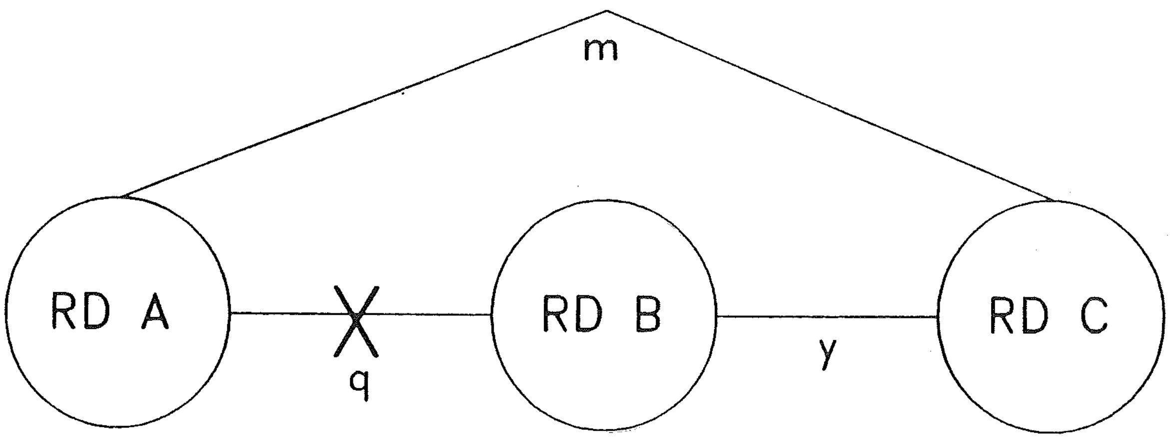
DRAWBACKS TO ROUTING DOMAINS

There are some drawbacks to setting up routing domains, especially when Phase IV to Phase V transition strategies are considered.

First, there is no OSI standard for dynamic inter-domain routing. This protocol is under development and it will take many months, if not longer, before it will be available. For the present, then, all inter-domain routing is static, and must be manually configured and manually fixed if a line goes down. This means network managers (or their operators) are responsible for adding and deleting addresses from the address tables and fixing circuit problems - manually. The more connections a routing domain has, the more manually intensive maintenance and operations become. Compare this with DECnet Phase IV routing where the network automatically attempts to repair a circuit outage by using a fallback path if available, regardless of the complicated physical topology - even in the middle of the night!

A classic example of the headaches introduced by manual maintenance is shown in Figure 7, "The two-hop problem". In this simplified drawing, Routing Domains A, B, and C are connected in a line. Routing Domain A and C normally communicate through B using circuits q and y. Circuit m exists between A anc C as a backup. Now, assume circuit q fails. Routing domain A recognizes q has failed, and begins to re-route its traffic destined for B and C over circuit m. So far, so good. To simplify this a little, let's look in particular at messages from A to C. A packet from A arrives in C over circuit m. C receives the packet and sends it to its destination in C. In response, the destination system tries to reply to the source system in A by sending a packet back into the network. However, because circuit q is not in C's routing domain, C has no knowledge of its failure. Therefore, C dutifully sends the reply packet to A *through circuit y*. B gets the packet and says "nope, I can't forward this, because circuit q is down" and sends it back to C. C gets the packet back, and again tries to send it to A *over circuit y*. C is really stupid about all this, but that's what static links can do to a network. C will never automatically re-route the packet over m, because C is never told that circuit q is

"The Two-Hop Problem"



- Normal static paths are RD B and C reachable through link "q" and link "y"."
- Link "m" designed as backup connection for A to C communication.
- Link "q" fails. *No* auto failover to backup link "m"
- Manual intervention required to re-route packets from C–A and B–A through C.

Figure 7.

down and thus should readjust its own routing to compensate. The packet will bounce around between C and B until it reaches its maximum cost or visits, and then it disappears: this is the "black-hole" effect of static routing. To use circuit m, a network manager in C will have to manually adjust the circuit parameters.

Second, network management cannot set a routing domain between two sites which use the same Phase IV area and must maintain Phase IV connectivity. This constraint is certainly the most restrictive for planning routing domain boundaries during the transition.

CONSIDERATIONS FOR THE DECNET INTERNET

Politics implies lots of routing domains. It is naive to assume that individual facilities, when having the ability to shield their networks, will not take the opportunity to do so. In the long run, setting routing domain boundaries will provide a mechanism for protecting a network's routing functions from problems in a neighboring routing domain. This means routing domains undoubtedly will be implemented down to site level, eventually.

However, prudence, responsive network routing, and preservation of Phase IV connectivity and network manager sanity indicate the network should support very few routing domains, at least at the start of the transition. It has been proposed that a logical place to set a routing domain boundary at the start of the transition would be across the Atlantic, between U.S. and European sites.

So, it is clear that we must eventually allow for the existence of many routing domains, but it is also clear that we will divide the DECnet Internet into only a few, and possibly just two, routing domains at the start of the transition. Therefore, the global transition strategy must incorporate mechanisms for identifying logical placement of new routing domain boundaries and coordinating the setting of these boundaries throughout the transition process.

SUMMARY

The need for a Phase V/OSI transition is clear. The limits of DECnet Phase IV protocols have been reached, and the Government is requiring implementation of OSI protocols for its agencies and departments.

The major issues for moving the Phase IV network to OSI/Phase V are being tackled for the DECnet Internet by the HSDCG. The implementation of addressing and naming are largely understood and accepted. A choice for the global "Phase IV prefix" still has to be made. The emplacement of physical name servers and the operation of the namespace in the Phase IV network is progressing.

More work remains to be done in planning for the use of routing domains in the DECnet Internet during the transition.

The general strategy to move the DECnet Internet Phase IV network to a Phase V network is to use an area-by-area transition plan, starting with network backbones, while preserving Phase IV connectivity throughout transition.

Detailed transition plans are being developed by the individual network participants taking into account the issues being coordinated by the HSDCG.

References:

1. Digital Equipment Corporation, DECnet/OSI Phase V, making the transition from Phase IV, Order no EK-DNAPV-GD, 1989.

2. ISO/TC 97 Information processing systems - Data Communications - Network service definition Addendum 2: Network layer addressing, ISO 8348 Addendum 2, 1988.

3. R. Kevin Oberman, EDWG DNANS Naming Policy and Guidelines, Version 1.2, May 1990, Lawrence Livermore National Laboratory

4. D. Oran, Intermediate system to intermediate system intra-domain routeing exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473), DP 10589, 1990.

5. U.S. Government Open Systems Interconnection Profile (GOSIP), Draft, Version 2.0, FIPS 146, April 1989