An investigation into some critical computer networking parameters

Internet addressing and routing

THESIS

Submitted in fulfilment of the requirements

for the Degree of

MASTER OF SCIENCE

of Rhodes University

by

Edwin David Isted

Department of Computer Science

December 1995

Abstract

This thesis describes the evaluation of several proposals suggested as replacements for the current Internet's TCP/IP protocol suite. The emphasis of this thesis is on how the proposals solve the current routing and addressing problems associated with the Internet.

The addressing problem is found to be related to address space depletion, and the routing problem related to excessive routing costs. The evaluation is performed based on criteria selected for their applicability as future Internet design criteria.

All the protocols are evaluated using the above-mentioned criteria. It is concluded that the most suitable addressing mechanism is an expandable multi-level format, with a logical separation of location and host identification information. Similarly, the most suitable network representation technique is found to be an unrestricted hierarchical structure which uses a suitable abstraction mechanism. It is further found that these two solutions could adequately solve the existing addressing and routing problems and allow substantial growth of the Internet.

Acknowledgements

I would like to thank my supervisors, Peter Clayton and John Ebden for their guidance and encouragement during the course of this project. Their enthusiasm during the course of this project and willingness to help was greatly appreciated.

Special thanks go to Greg Watkins and Tony Booth for their help in proof reading this document and providing valuable feedback.

Finally to all my friends and family who throughout the course of the last two years have provided their support and ideas.

Trademark information
BSD UNIX is a trademark of Berkley University

Foreword

Thesis Structure

This thesis is structured as three logical parts. The first part provides the background to the problem domain (Chapters 2 - 4). The second part deals with more high level protocol design issues (Chapters 5 and 6), and the last part is the actual protocol analysis and evaluation (Chapter 7 and 8).

The contents of each of these logical sections can be summarised as follows:

Part one: Chapter 2 gives a brief background to the Internet, its current growth trends, and usage. The primary problems that this thesis aims to examine are explained in chapter 3. A motivation for the work and a summary and evaluation of existing research into the area are provided in chapter 4.

Part two: Chapters 5 and 6 discuss higher level protocol design criteria. Chapter 5 deals with the design criteria for the existing IP protocol, whereas chapter 6 contains the criteria that the author believes should apply to the next generation IP candidate. The criteria put forward in chapter 6 have been selected after extensive research into general protocol design as well as some analysis of the Internet and what its current trends and future growth areas are likely to require from IPng.

Part three: Chapter 7 contains a technical analysis of 10 of the selected protocols and provides an analysis of how each of these protocols would affect the problems identified in chapter 3. Finally, chapter 8 gives an overall summary of the protocols with some suggestions for the IPng and for further research directions.

Thesis Referencing

RFCs (Request For Comments) have been used extensively in this research. Had this not been the case, the research in its present form could not have been conducted due to the rapidly developing

nature of the resource material used. The RFCs are published electronically, with a complete set available from any of the following sites :

> http://www.internic.net/ds/rfc-index.html
>
> http://andrew2.andrew.cmu.edu/rfc/rfc-index.html
>
> http://ftp.sun.ac.za/pub/rfc

The validation process for these RFCs is available in RFC 1800.

Document Layout KEY

1 Chapter Numbering
1.1 Subsection level 1
1.1.1 Subsection level 2
      Subsection level 3
        + Subsection 4
            ♦ Subsection level 5

- Point

Table of Contents

List of figures

L Introduction

The Internet is synonymous with the communication of ideas, E-mail and other electronic data. This interconnection of inhomogeneous networks, which evolved from its inception when it was still known as the ARPANET, is now the giant, globally spanning Internet. Included in the Internet are the protocol suites which were originally used on the ARPANET. Amongst these, the most prevalent are the multi-layered Internet Protocol containing TCP/IP. With the inception of the original Internet the design specifications were required to cope for at most a few hundred interconnected networks and a few thousand hosts. Considering these design specifications TCP/IP has coped admirably with the huge growth in both networks and hosts. Connected networks alone are conservatively estimated to number well in excess of 50000 [MERIT, 95], whereas the number of hosts with connectivity can only be roughly estimated at approximately 27 Million. This tremendous growth is placing incredible strain on the current IP protocols and a number of very serious problems are developing.

This thesis evaluates a number of the proposals which have been put forward as successors to the current version of IP, with special regard to routing and addressing issues. The motivation behind this thesis is the realisation of the importance of the Internet in everyday communication and hence the need for it to be able to evolve successfully. The emphasis on addressing and routing issues is due to the fundamental role which they play in the current IP protocol. These two areas are also those which are currently experiencing the most serious problems in terms of resource exhaustion and inefficiency.

This thesis endeavours to simplify the choice of a new IP protocol by evaluating existing protocols and providing some basic guidelines for addressing-related and routing-related issues.

The above view concerning Internet growth is reinforced by the quote *"A complicated mix of factors such as technological advances, political alliances, and service supply and demand economics will determine how the internetwork will change with time."* [Chiappa, 95]

2. Background

2.1 The Internet history

The Internet has developed during the third stage of what Corner describes as the three stages in network development and research. The first stage is characterised by the question *"How can we transmit bits across a communication medium efficiently and reliably?",* the second stage addresses *"How can we transmit packets across a communication medium efficiently and reliably?",* and the last stage is interested in *"How can we provide communication services across a series of interconnected networks?"* [Corner, 91]

Government agencies realised the importance and benefits of having interconnected networks during the early 70s. The initial research by DARPA *(Defence Advanced Research Projects Agency)* resulted in the creation of the flagship packet-switching network, ARPANET [Kahn, 94]. Early extensions to the ARPANET provided packet-switching interconnection between the already developed packet radio network (PRNET)(dual rate 400/100kbps spread spectrum radio) and satellite network (SATNET)(64kbps shared channel on Intelsat IV). The growth in network diversity was one of the main contributing factors which forced DARPA to investigate internetworking. The original internetworking architecture was developed by Vinton Cerf and Robert Kahn, who developed it from 1974 to 1978, during which time it went through four successive revisions. The fourth revision was the version which was eventually standardised, and adopted by researchers and academia. This internetworking protocol was later to become known as the TCP/IP protocol [Kahn, 94].

 The first true Internet networks using TCP/IP emerged in the early eighties when DARPA started converting the protocols on some of their research networks to use TCP/IP. By the end of January 1983 the transition of all computers on the ARPANET to use TCP/IP was complete, and connections onto these networks required TCP/IP compliance [Corner, 91]. Up to this time there had been a number of different standards used on the ARPANET.

To encourage the widespread acceptance and use of TCP/IP, DARPA funded BBN (Bolt, Beranek and Newman) to implement the TCP/IP protocols for the already popular BSD UNIX. Berkley University was then asked to integrate these modifications into their distribution of BSD UNIX. This put TCP/IP into the hands of 90% of computer science departments in the United States of America. The new software, together with the growth in use of computers and LANs, resulted in the TCP/IP protocol spreading to researchers in many other disciplines. Because TCP/IP implementations looked very similar to the existing UNIX routines, TCP/IP became very easy to learn and use.

In the early eighties the National Science Foundation realised that networking would be a crucial part of scientific research and started funding the CSNET, which linked ARPANET and all the NSF supercomputer centres in the USA. The new sites were placed onto the ARPANET by using a commercial network (Telenet) which gatewayed these sites onto the ARPANET {Kahn, 94]. At the same time other "community networks" were being established. The most notable computer orientated research programs which were established were DOE *(Department of Energy)* and NASA *(National Aeronautics and Space Administration).* Following the CSNET effort, NSF and ARPA continued working together to expand the number of users on the ARPANET. This expansion was severely constrained by the defence department's restrictions on the use of its network. This was the case until the mid-eighties when network connectivity had become sufficiently central to the functioning of the computer science community for NSF to become interested in broadening the use of networks to other scientific communities. The NSF supercomputers provided substantial motivation for the broadening of networks by providing limited access to these facilities via the ARPANET. These motivations, as well as ARPA's decision to phase out network research, led NSF to make the decision that it would assume leadership in the network development arena. As a result of this decision the NSFNET advanced network developed, linking all the NSF supercomputer centres with very high speed links (1.5Mbps) and providing access to the members of the academic community and to one another. Following the decommissioning of the last ARPANET node in 1990, NSFNET and a collection of smaller regional networks made up the core of the Internet in the United States. Other networks which developed in parallel with ARPANET during the course of the early eighties were

DECNET (Digital Equipment's network), SPAN (Space Physics Analysis Network), HEPNET (High Energy Physics Net), ESNET (Energy Sciences Network), and NSI (NASA Science Internet). These networks, as well as a number of other networks including numerous X.25 networks, were influential in the development of a multi-protocol networking environment which was subsequently embraced by the Internet.

Commercialisation of the Internet has a history which goes back to the late 1980s. The initial step was the limited access made available to commercial users by allowing commercial E-mail providers to use the NSFNET backbone. This was to enable communication between authorised users of NSFNET and the other federal research networks in the Internet at the time. The regional networks initially established to serve the academic community had already taken on non-academic customers in an effort to make themselves self-sufficient. As NSFNET's acceptable use policy limited backbone access to academic use, two privately funded and competing commercial carriers emerged. They were both spin-offs of government programs. UUNET emerged from the DOD-funded seismic research facility, and PSI (Performance Systems International) emerged from NYSERNET which operated in the New York and lower New England regions.

From its humble beginnings as a government research project the Internet has grown to become a major component of all network infrastructure, linking millions of users all around the world. Currently very little of the Internet is owned by any one person or organisation, neither is it controlled by any particular body. Most of the Internet is funded either from private sources or indirectly funded through the educational community. The trend towards commercialisation is continuing as more Internet service providers are providing Internet connectivity on a commercial basis [Kahn, 94].

2.2 The Internet growth

 Growth of the Internet in its earlier years was limited by the strict regulations regarding its use, as specified by the US defence which then owned and ran most of it. The first major growth occurred when the Internet was made available to the academic community in the early eighties.

From that point onwards, the growth has not slowed. Contributing factors to this growth are numerous, most notably the continued influences of NSFNET, the availability of LANs and commercial workstations, and also the current trend towards commercialising the Internet.

This growth was further encouraged by NSF, which stimulated and funded extensive improvements in routing technology and encouraged the distribution of networking software from its own supercomputing centres.

During the earlier years in the Internet history, computing was generally limited to mainframes and terminals, serviced by terminal servers. This technology did not encourage a huge growth in Internet expansion as this type of hardware was expensive and generally limited to a small number of organisations. The advent of the PC, LAN, and personal workstation era did away with this limitation, and now computing power could be distributed to a far greater audience, at far lower prices. Coupled with the availability of LANs came the awareness of the Internet and consequently increased growth of the Internet.

Commercialisation of the Internet is the current primary growth motivator, as the Internet, which once was the realm of research and academia only, is opened up to the wider world. This commercialisation is resulting in a very large number of commercial users and networks becoming part of the Internet. The effects of commercialisation can be clearly seen when looking at the growth figures since the beginning of the 90s when this trend first started.

Future markets for the Internet are predicted in the fields of entertainment, device control, and education. All these markets have the potential of being very large and developing extremely quickly.

2.2.1 Internet growth and its significance

To understand the significance of the growth of the Internet one must have some idea of the initial design parameters and how they have been exceeded. The initial design parameters in the Internet called for the interconnection of at most a few thousand hosts and a few hundred networks [Kahn, 94]. In practice, the growth of the Internet has been astounding. Not only has this growth been in the form of growth in the number of hosts and domains, but there has also been a tremendous growth in the number of services offered and used on the Internet. If one considers the growth figures since the split of the then ARPANET into MILNET and ARPANET (later to become INTERNET), it is evident that very few planners could have foreseen the tremendous growth. Figures for growth in the number of hosts can be seen in **Figure 2** [Lottor, 92]. These figures have been gathered using the ZONE program and reflect the data as collected by Network Wizards. Similar growth figures for the number of networks can be seen in **Figure** 3 [Lottor, 92]. The growth is so substantial that on average a network joins the Internet every 20 minutes.
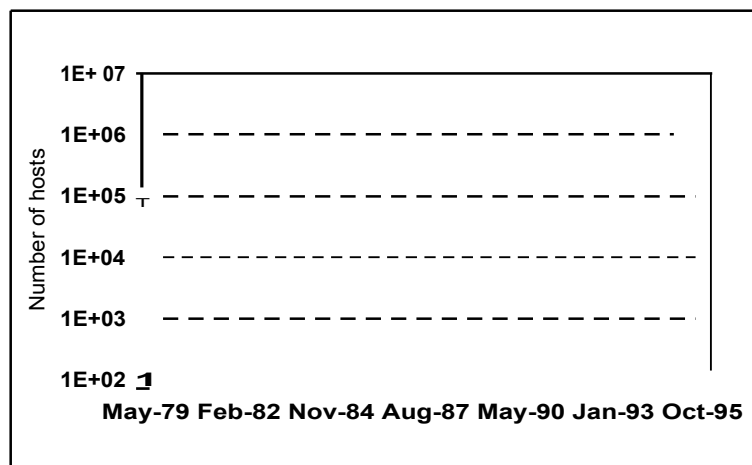
Figure 2 : Internet host growth chart
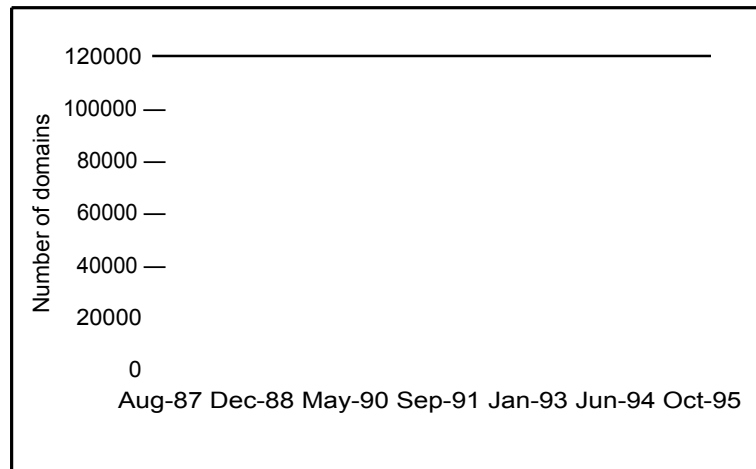
Figure 3 : Internet domain growth chart

When looking at the growth in the number of domains one needs to consider how this growth has been spread across the various address classes. **Figure 4, Figure 5 and Figure 6** give some sort of indication as to the growth within the various address classes (address classes will be discussed in chapter 3). Class A addresses have effectively levelled off and they are static at the moment. This may be as a result of the extreme reluctance of the InterNIC to allocate any class A addresses.

Class B addresses, on the other hand, show the alarming trend which has become a primary motivating factor for a replacement for IPv4. Considering that the limit on the number of
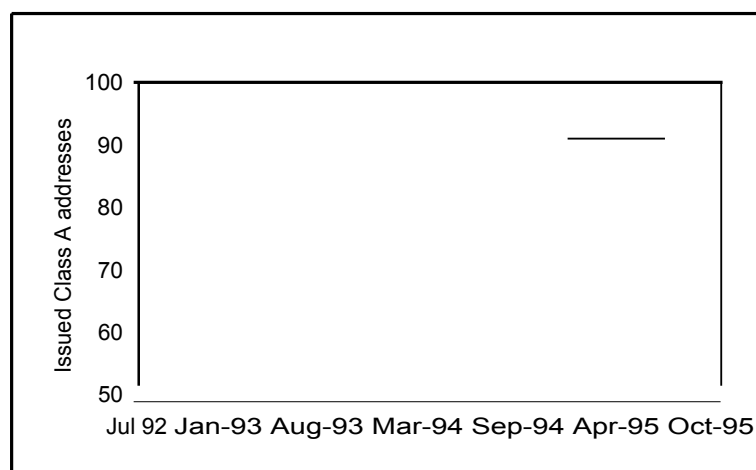
Figure 4 : Growth of Class A networks in the Internet

allocatable class B addresses is approximately 2^{15} and taking the current growth rate of approximately 25% per year, there is a very high probability of depletion of the class B address space within the next few years.
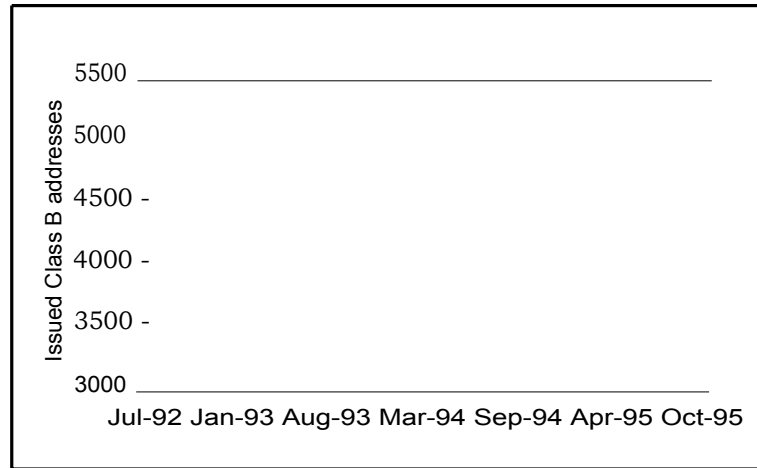
Figure 5 : Growth of Class B networks in the Internet
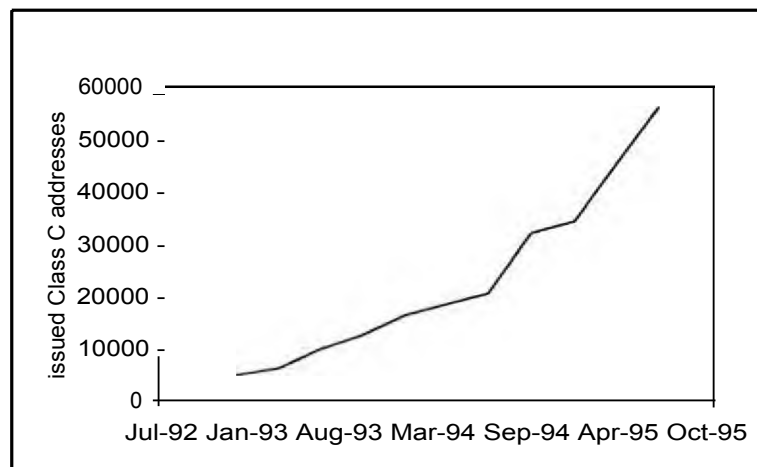
Figure 6 : Growth of Class C networks in the Internet

In an attempt to slow the growth of class B addresses, Classless Interdomain Routing (CIDR) [Fuller *et al,* 93] has been implemented at the inter-domain level in the routing hierarchy. This scheme involves allocating a number of class C addresses to smaller organisations instead of providing them with a class B address which they would not be able to utilise fully. The effect of implementing CIDR can be seen clearly when looking at the growth rate for class C addresses between January 1995 and July 1995. Over this period the number of allocated class C addresses

grew by a staggering 64%. This trend is very likely to continue as the overall growth in the Internet continues, and more individuals become "Internet-connected".

The tremendous growth in the Internet over the past two decades has resulted in some severe problems, many of which may be directly linked to fundamental design flaws in the architecture.

2.2.2 Problems in measuring the growth of the Internet

The ZONE (Zealot of Name Edification) program walks through DNS trees to build a host table. Prior to the extensive use of DNS tables (1981-1986), the statistical data was gathered from the central SRI-NIC host tables, in which every host was registered. Once the DNS method was implemented in the Internet (completely implemented from 1988 onwards), statistical data collection became a little more difficult due to a number of bugs within the software used by the ZONE program. Some sites refuse zone transfers or are not registered in a domain name server. The number of sites which refuse zone transfers is relatively large at around 800 out of 17000 domains [Lottor, 92]. It can thus be said that all the figures represented in the data given, represent the minimum number of hosts and domains within the Internet.

A large number of sites use techniques which prevent the ZONE system from being able to gather data from their sites. Some of these techniques include using mail gateways, firewalls or mail - forwarding techniques.

The last problem is one of scale. Since the Internet has grown to be such a large and distributed entity, the cost of collecting the growth data is becoming very expensive, not only in time, but also in pure processing and communication costs.

2.2.3 Growth measurements hi isolation - the problems

Since the inception of the Internet, it has been undergoing a constant cycle of growth and subsequent refinement in all associated aspects. The initial hardware which was used in the core

gateways was soon replaced by more powerful hardware, which had dedicated software and hardware to handle the task of providing adequate packet forwarding. Similarly, the physical links between sites have been - and are still being - upgraded at regular intervals. Some of the initial lines operated at 1900 baud, whereas most of the core lines today are at least TI' if not already T3[2] lines. Refinements to the IP protocol have also been made along the way. Although these refinements may not appear spectacular, they too have had a tremendous effect on the performance, usefulness and useability of the current Internet.

2.3 The Internet usage

The predecessor of the Internet, namely the ARPANET, evolved from research efforts within the ARPA defence contractor. ARPA were commissioned in the early 70s to investigate packet switched inter-computer networking. Essentially the only purpose of the Internet at this time was for research purposes, directly related to military requirements. The types of experiments which were run on these test-bed networks were large-scale scientific time-sharing systems, where users logged into mainframes via terminals. All network research up to this point was predominantly government funded.

During the course of the early 80s the usage diversified into the field of computer science. Eventually the extensive computer science research done using the existing networks prompted the formation of CSNET, which later became NSFNET. With the spread of TCP/IP, research using the Internet began to extend beyond the immediate computing related fields.

The first signs of commercialisation of the Internet occurred in the late 80s when commercial E-mail providers were allowed to use the NSFNET backbone for a limited class of their E-mail. Soon many of the smaller regional networks began accepting non-academic customers. By 1990 there was significant commercial growth, this was further accelerated by the provision of better

---

[1]T1 lines - T1 lines are carriers which have a data rate of 1536000 bits/second [Stevens, 94].

[2]T3 lines - T3 lines are carriers which have a data rate of 44210000 bits/second [Stevens, 94].

routing algorithms and the development of a number of CIXs (Commercial Internet Exchanges) which allowed for the interconnection of commercial networks to other networks [Kahn, 94].

2.4 Conclusion

The historical background to the Internet has been a checkered one, starting out as a pure research experiment and then being slowly turned into a useful medium for communication. Today, the Internet can be seen to be a major medium for the transfer of information in our modern society. The growth in the use of the Internet has been accompanied by a similar and perhaps far more important growth in the technology being used in the Internet. This falls into the three predominant areas of hardware, software, and physical links. Accompanying the Internet's growth, its usage is encompassing far more diverse fields as it is becoming more accessible to a larger portion of the world's population.

3. The Internet and its problems

The tremendous growth of the Internet over the past two decades has resulted in some severe problems, many of which may be directly linked to fundamental design flaws in the architecture.

TCP/IP, which was designed in the early 70s, was thought to have a sufficiently large address space with 32 bits allocated for the **IP** address pool. This address pool was then split into distinct address classes (Class A, B and C), each class having a predefined number of bits for indicating

		High order bits	Low order bits	
Class	**Class prefix**	**Net**		**Host**
A	0	7 bits		24 bits
B	10	14 bits		16 bits
C	110	21 bits		8 bits
	111	*extended addressing mode*		

Figure 7 : IPv4 address classes and bit allocation

the network number, and the rest for host numbers within that network. This is shown in **Figure 7. The** use of three distinct address classes was perceived as providing sufficient address aggregation[3] for the foreseen growth and hence a reasonable routing hierarchy for the **TCP/IP** protocol. See **Figure 8** for more details as to how the number of routes advertised within a routing table can be estimated.

---

[3] Aggregation - grouping of addresses which share some characteristic, typically in IPv4 that meant the same network number.

Core routing Table	Local domain (Secondary) routing Table
Entries in table	
Entries for all network prefixes	Entries for local network prefixes
Optimisations	
none - no default routing	uses default routing for unknown routes
CIDR - Single prefix advertised for networks which have been aggregated on network number .	CIDR - Single prefix advertised for sub-networks which have been aggregated on network number.
Number of possible entries:	
$\sum$ALL ALLOCATED NETWORKS $_{max}$ $\geq 2^7 + 2^{14} + 2^{21}$	$_{(LOCAL}$ ALLOCATED NETWORKS + DEFAULT ROUTES)

Figure 8 : Routing table entries using IPv4 address classes

In hindsight, these two design decisions have proved to be the most serious flaws within the current IP protocol. Despite the two problems being seemingly unrelated, they have to be considered together due to their complex interaction. This interaction will be discussed shortly in section 3.3. First we discuss the two major problems and exactly what each of them means:

3.1 Addressing problems

- Internet will run out of class B addresses

- Internet will eventually run out of 32 bit address space

3.2 Routing problems

- Routing is fast becoming impossible as the number of IP network numbers grows beyond the point where they can be routed efficiently by the current routing algorithms and hardware. (Routing tables in core routers can not cater for the number of networks which have to be explicitly advertised.)

- Type Of Service considerations and strict policy routing are not supported by the current Internet routing policies.

3.3 The interrelation between routing and addressing

To understand the interrelation between the routing and addressing problems on the Internet, we have first to look briefly at the terms used in networking. Of particular concern are names, addresses (locators) , routes, and selectors.

- **Names :** names are unique within some context; they say nothing about an object's location or how to get there.

- **Locators (addresses) :** locators give information which can be used to perform a lookup which will provide routing information. Locators belong to the interface and name the interface where you are connected to the network. This means that locators change as physical host locations change.

- **Route :** routes provide a mechanism for getting from point A to point **B.**

- **Selectors :** selectors are used by the forwarding mechanism to determine how the packet is going to be forwarded.

With **IPv4 the IP** numbers are assigned from a 32 bit address space. However, the **IP** number is overloaded and is used to represent a number of things for each host. In the first place it is used to uniquely identify the host (i.e. a name), and at the same time is also being used for routing decisions and hence is also used as a selector. The proper classification for an IP number is a locator (address).

There is general agreement that we have to move to an addressing space which allows for better aggregation of addresses. Address aggregation is defined as groups of addresses which share some underlying characteristic, be that an AD(Administrative Domain) or Geographic location [Clark *et al,* 91]. This aggregation will hopefully accomplish a number of things. In particular it will specify a domain in which a certain policy is implemented with regards to routing elements,

network management and/or TYPE OF SERVICE provision. At the same time this aggregation should provide some method for efficient routing.

Historically, the Internet address space was divided up into Classes, with each class providing a degree of address aggregation. Each class supports a finite number of hosts, this number being dependant on the particular class: i.e. Class A supports 2^{24} hosts (2^7 nets), Class B 2^{16} hosts (2^{14} nets) and Class C 2^8 hosts (2^{21} nets) [Postel, 81a]. The idea behind this classification was that each organisation be given an address appropriate to its size (with some lee-way for future growth). Unfortunately most organisations are too big for Class C, and so are issued with a Class B, which they then utilise very poorly. This allocation of poorly utilised addresses puts a large strain on the Class B addressing space.

There have been a number of temporary solutions suggested to slow down this growth in the Class B category. One of the solutions, Classless Inter Domain Routing (CIDR), does this and instead shifts some of the growth into the Class C category. CIDR, which is already in use in certain domains, ignores the old address classes and instead does a form of bit masking to achieve more efficient routing [Rekhter *et al,* 93a].

3.4 Proposed solutions

Proposals as to how to tackle the problem of address space depletion fall into three distinct categories;

- Expand the address space

- Reuse the current address space in a local domain

- Keep the current address space and perform some sort of address mapping/hashing [Clark *et al.,* 91]

All the proposals for IPng[4] deal with the address space issue in one of the above mentioned ways. The relative advantages and disadvantages of each will be discussed in more detail when the actual proposals are dealt with in Chapters 7 and 8.

3.5 Conclusion

Simple conversion of the address space to whatever solution is finally adopted is not enough; the final address space must also take into account the issue of providing a more scalable routing architecture. Coupled with this goal of more efficient and scalable routing comes a move, from our model where our packet forwarding units (gateways) are stateless, to one where they contain at least some state information related to each through-connection.

The most immediate and threatening problem with the IPv4 protocol is the issue of problematic routing due to excessively large routing tables in core routers. This problem exists already, whereas the final depletion of address space is in fact a future problem which has net yet been encountered. It is imperative that any IPng tackle both these problems with equal creativity and that neither gets relegated to take a back seat in IPng.

---

[4] IPng - Internet Protocol next generation. This is the generic name given to the next IP protocol.

4. Motivation for this project

4.1 Introduction

This chapter contains a summary of the related research in the area. The other research done into the area of Internet protocol problems by [Dixon, 93] and [Clark *et al,* 91] will be outlined briefly. A further section in this chapter will indicate the uniqueness of this thesis and motivate for the approach which has been taken in creating this work.

4.2 Related Research

At least two other authors have examined the area of the future IP protocols and their required characteristics.

The first work by [Dixon, 93] is a generic discussion as to why the existing IP protocol is inadequate, and gives three IP proposals a very brief overview. The proposals mentioned in this work are PIP, SIP and TUBA. The document produced by Dixon can only be described as a very cursory discussion document, as it does not provide any insight into the operation of any of the protocols introduced. The discussion on the existing inadequacies of IPv4 is somewhat more useful, in that it outlines an approach for the design of the next IP protocol.

A second work in the area by [Clark *et al,* 91] is more substantial than that of [Dixon, 93], yet also only deals with the future requirements of a new IP protocol. No actual protocols are discussed at all in this document, with the emphasis being entirely focussed on 5 areas of architectural evolution.

The five areas which are focussed on by [Clark *et al,* 91] are the following:

- Routing and addressing

- Multi-protocol architecture

- Security

- Traffic Control

- Advanced applications

Each of these areas is introduced briefly, the problems outlined and a suggested approach given for solutions.

43 Motivation and approach of this thesis

The motivation behind this thesis is to investigate two areas of the Internet architecture, namely routing and addressing. Routing and addressing are interrelated, and are absolutely fundamental to the functioning of any network protocol, and so too for the functioning of the Internet. The aim of this thesis is to establish a set of requirements for the future Internet routing and addressing, and then to evaluate a number of the proposed protocols against these requirements or criteria.

43.1 Motivation for the criteria selection

❑**Broad Criteria outline**

The criteria have been chosen based on the work done by a number of authors in the associated fields of research. The work done by [Clark *et al,* 91] and [Dixon, 93] provided a broad skeleton and a starting point for these criteria.

The broad outline criteria relate to issues such as protocol migration, backward compatibility, accountability, data integrity, and security. Protocol migration is a very necessary consideration if the new IP protocol is to successfully replace IPv4. Coupled

with this migration is the necessity for providing inter-protocol communication, at least during the protocol migration stage. Three criteria which can be used as motivating factors for migration are: data integrity, security, and accountability. These three areas are of importance especially with the current commercialization trends of the Internet. All of these criteria are explained in more detail in sections 6.1 and 6.3.

❑**Addressing criteria**

Address related criteria are based on work by Chiappa (**IP** addressing requirements), Tsuchiya (subnet number assignment), Gerich (management of **IP** space) and Gross (IESG deliberations for routing and addressing) [Chiappa, 94b][Tsuchiya, 91b][Gerich, 93] [Gross *et al,* 92].

The two main addressing criteria of importance to this thesis are: the address pool size and the network pool size. The address pool size indicates the number of hosts which can be accommodated by a given protocol, whereas the network pool size indicates the number of networks which a given protocol can support. Both of these criteria are critical to the new **IP** protocol if it is to cater for the future growth of the Internet. These criteria are dealt with in more detail in section 6.3.

❑**Routing criteria**

Criteria for routing issues and advanced routing mechanisms are based on work by Saunders (routing services), Simpson (mobility), Teraoka (mobility considerations and dynamic reconfiguration) and loannisdis, Kleinrock and Aziz (routing mechanisms to support mobility) [Saunders, 95] [Teraoka *et al,* 94] [Simpson, 94] [Ioannisdis, 94] [Kleinrock, 94] [Aziz, 93].

The routing hierarchy forms the basis for all routing services within networks. It also determines to a large extent the efficiency of information distribution and delivery. Hence,

this criteria forms an indispensable part of this thesis. The main emphasis related to this criteria is the protocol's ability to route datagrams over a very large and dynamically changing internetwork. Further aspects relating to the routing hierarchy are explained in section 6.3.3.

]Growth related criteria

Further criteria relating to Internet growth and market related issues are based on work by Curran [Curran, 94] and Kahn [Kahn, 94a].

Growth related criteria considered in this thesis are issues such as support for advanced applications, advanced network services, network management, support for a multi-protocol architecture, and long term connectivity. Growth within the Internet cannot continue at its current rate if more efficient and powerful network management tools are not developed. The successor to IPv4 has to ensure that resource management requires minimal human intervention and can cater for advanced applications which may require specialized network services in a multi-protocol architecture. Ultimately long term connectivity has to be ensured if the Internet in its current form is to continue. The above-mentioned criteria are dealt with in more detail in chapter 6.

4.3.2 Differences between this project and previous work

This project is different from previous research in that it deals not only with the design considerations for the next **IP** protocol but also evaluates existing proposals against these requirements (with respect to addressing and routing). Furthermore, it considers a wide selection of current proposals, which permits a more objective view of the solutions available.

4.4 Conclusion

This thesis takes a fresh look at the specific problem of addressing and routing, which is

fundamental to the entire Internet protocol suite, and investigates possible solutions.

5. Specification Requirements

When looking at the specification requirements for the IPng protocol, they must be looked at in conjunction with the specification requirements for the already operating IPv4. The reason that IPv4 has to be considered is that IPng is being implemented to gradually phase out IPv4. In a sense IPng is a protocol upgrade for the Internet.

It will become evident when reading the list of IPv4 design parameters that the degree to which the design parameters have been complied with varies for each parameter specification. The three primary specifications have been met very well, whereas the remaining four have only been partially satisfied. An explanation for this anomaly could be the environment in which IPv4 was developed and initially used [Postel, 81a] [Kahn, 94]. A shift in emphasis of the original IPv4 design parameters is foreseen as the Internet becomes a more commercialised internetwork, gradually moving away from the strictly military and academic network where it first developed.

5.1 The DARPA Internet Protocol Requirements

The development of a packet switched datagram network was initiated by the Advanced Research Projects Agency of the US Department of Defence. Included within this organisation's development efforts was the Internet Protocol (IP) and the Transmission Control Protocol (TCP). Despite the definite design parameters laid down for the Internet Protocol, the overall design philosophy behind the TCP/IP protocol has evolved with time. Evidence of this is the fact that something as fundamental as the IP TCP layering was only a later addition to the protocol.

The primary objective of the DARPA Internet Architecture was to *"develop an effective technique for multiplexed utilization of existing interconnected networks"* [Clark, 88]. Fundamentally, the components of the Internet are networks which are connected together to provide some larger set of services or functionality. During the initial stages connectivity between the ARPANET and the ARPA packet radio network was sought. With time this goal became far

more general and it was assumed that a general inhomogeneous set of networks would eventually be connected into the system (Internet).

The multiplexing technique chosen, namely packet switching was selected because of its suitability to the applications being supported, such as remote login. The fundamental nature of the traffic was to be datagrams which further suited packet switching technology.

The above choice gives one the fundamental structure of the Internet: " **a** *packet switching communication facility in which a number of distinguishable networks are connected together using packet communications processors called gateways which implement a store and forward packet forwarding algorithm."* [Clark, 88]

More specific design parameters relating to the effective interconnection of networks were then specified by a list of prioritized goals which the DARPA Internet had to satisfy, namely:

  *Reliability   *(Primary Specification)*

- **Multiple** Service Types   *(Primary Specification)*

- Variety of networks   *(Primary Specification)*

- Distributed resource management   *(Secondary Specification)*

- Cost effectiveness   *(Secondary Specification)*

- Simple host attachment   *(Secondary Specification)*

  *Accountability   *(Secondary Specification)*

These will now each be examined in more detail.

5.1.1 Reliability

Communication must be capable of continuing despite loss of networks or gateways. This matter of survivability comes from the fact that the architecture was designed to be used in a possibly hostile environment, where intermediate entities within the internetwork are not guaranteed to be stable or permanent [Clark, 88]. This goal can be expressed as follows;

> If two entities are communicating across the Internet, they must know nothing about the underlying temporary/permanent disruptions of service provision. Any underlying disruption must be catered for by the system being able to reconfigure itself to re-establish the communication "channel", without the higher level applications at either end having to reset their state information.

The automatic reconfiguration ability of lower levels within the internetwork provides only two possible states for the higher level applications, namely:

- connection

     or

- complete partition failure.

This level of service provision poses interesting questions related to the distribution of state information within the network. Allowing for "dynamic" reconfiguration means that state information cannot be kept at lower levels within the architecture, since this can easily be lost after a temporary failure.

Alternatively, the state information for these connections is stored in the intermediate packet switching nodes (or gateways). Storing state information in gateways poses further problems. The foremost is that, to protect from state loss, the information has to be reliably replicated using a robust replication algorithm. This would make the gateway software applications very difficult to design and build.

The final placement for state information was chosen as being within the entities which were

using the service. This principle, known as fate-sharing, says that it is acceptable to lose state information only if one of the entities using the service is lost at the same time as the loss of state information. Fate sharing has a number of advantages over replication, most importantly: fate sharing can protect against intermediate failures, whereas replication can only protect against a number of specific intermediate failures. It is also less expensive in terms of resources and time to engineer and implement. Fate sharing has two further implications;

- Firstly, the gateways may not contain any essential state information pertaining to on-going connections. They must in effect be stateless packet-switches. (i.e. datagram networking).

- Secondly, a larger degree of trust is placed on the host machine with respect to its algorithmic integrity, since the host machine's algorithms are required to cope with sequencing or data delivery failure.

5.1.2 Multiple Service Types

Support must be provided for multiple service types in the transport layer [Clark, 88]. Different service types are distinguished by differing requirements relating to speed, latency and reliability. Traditionally, the most common service would be a *bidirectional reliable data delivery service,* which is sometimes also known as a *virtual circuit.* This service type suits applications such as remote login and file transfer well, hence it was the first available service type on the Internet architecture. Soon it was realised that there were in fact a number of variations with regards to the higher level applications which used the TCP layer. Some, such as remote login, require a low delay and low bandwidth service, whereas others, such as remote file transfer, require a high bandwidth with delay being of less concern.

Initially the design philosophy behind TCP was that it be general enough to support any needed type of service. However, as the need for a wider range of services became evident, it was soon recognised that insisting that TCP provide support for all these services would make it

unnecessarily complex.

Examples of services which fall outside the domain of TCP are the XNET Cross Internet-debugging protocol and the transmission of real-time digitised speech or video. Both of these service types placed requirements which were well outside the designs of the TCP protocol; for instance, the XNET debugging protocol required added complexity with respect to timers and connection status (which could include unreliable and out of order delivery, something TCP explicitly tries to prevent). Likewise, the transmission of digitised speech requires a service which need not necessarily be completely reliable, but which minimises the delay between packet delivery. In the case of audio transmission, providing reliable delivery could seriously retard the timeous reassembly of speech data in "real time", and the time lag due to the retransmissions of lost packets could easily cause glitches in the reassembly process [Clark, 88].

As the limitations of a single transport layer service were found, it was soon decided that more than one transport service be provided within the architecture, each being tolerant of other services within the same protocol layer. TCP was then intentionally redesigned to provide a *reliable sequenced data stream.* IP, on the other hand, was considered to be the basic building block from which a variety of services could be built, this building block being the datagram. Since datagram service is a "Best effort" delivery service, one can build both unreliable and reliable delivery services from it, by adding acknowledgements into the higher level transport layer.

A previously undiscovered problem with the provision of multiple service types came from the physical networks on which they operated. In most cases the physical network had been designed with some explicit support for an intended service. Precisely this support made it problematic for the implementation of other service types over the same medium.

5.1.3 Variety of Networks

The Internet architecture must accommodate a variety of networks. Paramount to the success and acceptability of the Internet is the fact that it can make use of a large number of different network technologies, including both commercial and military facilities. These include a large number of different network types, such as long-haul networks, local area nets, satellite links, serial links, high speed asynchronous connections, and many more.

To achieve this wide operating base, the Internet protocol makes some very simple assumptions with regard to the underlying network structure and the services which it can provide. These assumptions are that, " *the network can transport a packet of reasonable size, with a reasonable reliability using a suitable addressing mechanism across its physical medium."* [Clark, 88]

It is important to notice that a very large number of services are not explicitly assumed from the network, for instance; network level multicast or broadcast facilities, reliable and prioritised sequence delivery etc. The reason for this is that, had these services been included within the initial specification, any network requesting to join the Internet would have to support all these services at the network level, or there would be software required in the network interface to simulate these services from end to end of the particular network. Clearly this scenario was undesirable since it would require re-engineering the services for every single host interface to the network, hence the reason for providing these services within the transport layer instead.

5.1.4 Distributed Resource Management

Some sort of distributed resource management must be provided for within the Internet architecture [Clark, 88] [Clark *et al,* 91]. This goal has been partly met in so much as the Internet gateways are not managed by any one organisation, i.e. there is distributed resource management. It has fallen very short in the sphere of *sufficient tools* for distributed management especially in the field of *routing.* The existing routing rules are based on policies for resource usage. However, they are implemented in a very limited way, through the manual setting up of tables.

5.1.5 Cost-effectiveness

The Internet architecture must be cost-effective. Consider the case of a packet with an Internet header of approximately 40 bytes containing the data for a single keystroke during an interactive login session. This represents a huge overhead for the service being offered. Unfortunately this underutilisation of communication resources is very difficult to tailor in a environment which uses so many different interchange services. Another area of inefficiency is that of packet retransmissions. Since there is no attempt to recover lost packets at the network level, it is often necessary to retransmit packets across the entire path, which could be repeating the transmission through a large number of the intermediate networks.

There is an efficiency threshold associated with this retransmission cost. For low retransmission rates (<1%) the cost of the packet recovery code in the network layer is higher than merely retransmitting the lost packets from the end point hosts [Clark, 88].

5.1.6 Simple Host attachment

The Internet architecture must allow for simple host attachment. The cost associated with connecting a host to the Internet is somewhat higher than in many other networks, due to all the mechanisms being provided on the host for all the different services [Clark, 88]. All the services and mechanisms such as acknowledgement and retransmissions strategies mean an increase in complexity. It is as a direct result of the design decision made within the protocol that this is the case. Unfortunately this also means that if there is a poor implementation of the protocol on a host it may damage the network as well as the host.

5.1.7 Accountability

There must be accountability for the resources used in the Internet architecture. Unlike a commercial network where this goal may have been considerably higher, there is very little in the way of support for accurate accountability of resource usage on the Internet. Network managers

have very few tools available which they can use to determine exactly what services are using their available physical medium the most, or which services are expensive in terms of processing time on particular network entities [Clark *et al,* 91].

A typical example would be a network manager wanting to find out the costs associated with operating a newsfeed through his/her network to a given number of satellite networks. Trying to compute this kind of cost with the current network tools is very difficult and at best the accuracy of the results are questionable.

5.2 The IPng Protocol requirements

Any replacement for IPv4 will have to assume an evolutionary rather than revolutionary approach to protocol development [Gross *et al,* 92]. The reasons behind this are twofold. Firstly, the existing user base for IPv4 is not going to accept a new protocol if it does not provide them with a reasonably seamless transition from their present protocol (Seamless in the sense that all their current applications and ways in which they make use of the Internet will have to remain fairly similar). Secondly, and perhaps more from a systems development perspective, the development of an entirely new protocol is not a trivial task and it would be foolish to ignore all the lessons which have been learnt through the evolution of IPv4, These lessons should be well remembered and applied to any replacement protocol. Certainly any successor to IPv4 will have to prove at least as successful as IPv4 if it is to succeed widely. Making use of the many design lessons learnt through IPv4 is one way to ensure this success.

To distill the essential protocol requirements for IPng one has to consider the motivation for this new protocol. The primary motivator in this case is the tremendous growth of the existing [P protocol. This growth is fast outstripping all predictions of "reasonable" network growth as foreseen when the protocol was first designed (See chapter 2). Not only is the use of the network increasing, but also the types of applications which are being used are becoming far more diverse [Carlson *et al,* 94] [Dixon, 93]. This growth in application diversity is another aspect which is placing considerable strain on the existing Internet protocol. Two areas where this problem is

clearly evident are in the fields of mobile computing and multimedia applications, both of which are solved sub-optimally in the present situation.

Clearly there are two primary design criteria which become evident from the above discussion. The first is the need to be able to sustain future network growth and the second is the need to provide some sort of IP protocol which makes it easier for application developers to better utilise the network resources for new applications. Stemming from the realisation that a new protocol is required, some very well designed transition mechanism is also required. The importance of this protocol transition mechanism should not be underestimated, as it determines to a large extent the success and early acceptance of any new protocol [Gross *et al,* 92]. Furthermore, this mechanism should be of such a nature that protocol upgrading can be incremental and occur without external intervention such as enforced flag days when certain upgrades have to occur. The essential design parameters for IPng can be summarised in **Figure 9.**
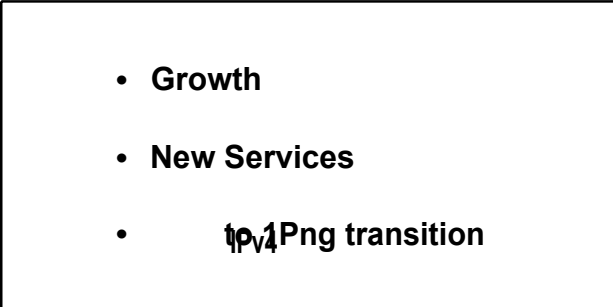
- **Growth**

- **New Services**

- **ipv4Png transition**

Figure 9 : *Critical* IPng design criteria

To produce a more defined picture of the exact requirements for IPng, we can look at each one of the three facets, i.e. Growth, New services, and Transitional mechanisms in more detail.

5.2.1 Growth

Areas which have to be addressed as a result of the tremendous growth in the Internet are the following: Routing Hierarchy and Address Space.

❑Routing Hierarchy :

Currently the routing hierarchy is insufficient to enable efficient routing across a very large Internet. The two tiered system which is in use has severe limitations, in that it requires very large routing tables on core routers, which makes them slow and prone to routing problems [Dixon, 93] [Gross *et al* , 92],

A replacement routing hierarchy will require a sufficiently layered hierarchy to provide efficient routing and also prevent routing overloading at any of the hierarchy levels [Gross *et al,* 92] [Hinden, 93].

DAddress Space :

In the euphemistic statement as presented by the Robert Hinden in his overview of the IP next generation protocol he says the following about IPng: "....*addressing and routing must be capable of handling reasonable scenarios offuture growth."* [Hinden, 95] Growth areas in the future are seen in the fields of the computing market, mobile communications market, entertainment market, and also the device control market. Presently the computing market is the only real market for the Internet. This market on its own has been able to sustain exponential growth over the last decade (See chapter 2). The other markets are seen as mass markets which have the potential for developing in parallel. Their combined influences on the Internet could potentially be enormous. Growth alone could increase by orders of magnitude (See section 22).

5.2.2 New Services

The issue of New Services on the Internet is multifaceted and motivated from two distinct directions. The primary need is driven by the commercialisation of the Internet [Leiner, 94] and hence the requirements for Quality Of Service issues and the associated security and routing considerations. Further pressure for new service types comes in the form of being able to provide

some strong motivating factors for users to migrate to the new protocol. These may be in the form of better support for new physical media (e.g. high speed broadband technologies), or more advanced option support (e.g. more efficient forwarding and more flexibility for introducing new options) [Clark *et al*, 91] [Curran, 94].

The Internet as developed since the early 70s has had a history of starting off as a military project and then moving into the more general realm of academic computing. Neither of these two environments place a very high emphasis on financial control, hence the almost complete lack of resource accountability. Other fields which have been partly ignored are those of cost effectiveness, simple host connection mechanisms and remote resource management. These aspects will have to become far more prevalent in IPng as the protocol moves into a far more commercialised arena [Curran, 94] [Clark *et al*, 91] [Hinden, 94a].

5.23 Transitional Mechanisms

The success and acceptance of IPng depends very largely on the transitional mechanisms which it provides [Hinden, 94a].

The most basic assumption which has to be made for transitional mechanisms is the continued operation of the Internet. This implies a number of things;

- There can be no disruption to any of the current IPv4 users of the Internet.

- Users of the new IPng should be provided with at least as much functionality from their first session onwards as the users using IPv4.

Continued operation of the Internet can only be partially guaranteed if IPng is in place long before the existing IPv4 Internet reaches its critical overload point. There are a number of important facts which have to be considered here. The IPv4 and IPng transitional phase will probably be one in which a very large number of hosts and gateways run dual protocol stacks, to provide some

sort of backward compatibility between IPng and IPv4. Currently many gateways are running at very high utilisation levels [Eidnes, 94] [IPng news group] and forcing them to run a dual protocol stack is only going to make their performance worse. Should this dual protocol stack only be implemented when the Internet is near the critical overload point, the gateways will more than likely not be capable of operating at all.

Little or no disruption to the users of IPv4 can only occur if all the elements within their Internet are either kept as is or replaced with fully functional and backwardly compatible components. Upgrade interdependencies must be kept to a minimum and a diffuse and incremental upgrade pattern must be supported. With the introduction of a completely new addressing scheme, there must be minimal overhead for processing existing IPv4 address in an IPng system.

Providing users of IPng with the basic network services can only be done if they support dual protocol stacks and are able to make of the existing IPv4 services as well as the new IPng services. Should IPng only be deployed once the IPv4 Internet has reached capacity, ensuring this kind of backward service compatibility could be very difficult as there would be a number of technical problems, not least of all being the lack of a suitable "temporary" 32 bit address for the IPng host to use when communicating with the IPv4 service provider [Hinden, 94a].

5.3 **Conclusion**

The design of the IPng protocol cannot be considered in isolation from that of IPv4, for the simple reason that IPng has to be seen as an upgrade of IPv4 and not a replacement. As will become apparent in the next chapter, IPng places a very large emphasis on providing at least the same kinds of services as IPv4, whilst at the same time providing complete interaction between the two versions. This is clearly a case of protocol evolution and not one of simple protocol replacement.

6. Design Issues : Operational

Three separate areas relating to the design issues for a new protocol can be identified: those which influence the protocol acceptance to its users, those reasons why a new protocol has to be adopted in the first place, and finally those underlying advantages which the new protocol presents to the more technical users and system administrators. These three areas have been labelled Phases for this discussion and will be treated separately, despite them all being integral parts of the new protocol.

6.1 Phase 1: Protocol acceptance issues

Protocol acceptance issues relate to how the protocol is going to be made more acceptable to its new users, or more importantly how users of another protocol are going to react to the new protocol when they are encouraged or forced to start using it. There are a number of protocol acceptance issues. Three important ones, namely migration, backward compatibility, and incremental changeover will be discussed in more detail.

6.1.1 Migration

Migration of users from IPv4 to IPng is a very important aspect to the success of IPng. Should a large number of users decide that IPng is not worth migrating to, the existing problems with IPv4 will only prevail and possibly even be exacerbated. It has been suggested that successful migration from IPv4 to IPng will depend on three core factors [Hinden, 95] [Curran, 94] , namely ;

- New functionality in **IPng**

- Reduced cost as a result of using IPng

- Connectivity to otherwise unreachable hosts

The importance of each of these factors has been estimated based on the current trends and likely future scenarios in the Internet (See chapter 2). As a result of this estimation it has been concluded that the cost factor would be the least important, followed by the connectivity issues. However, it must be pointed out that the connectivity issue is very strongly tied to the short term growth of the Internet.

The strongest motivating factor for migration from IPv4 to IPng is the added functionality which IPng should present [Eriksen, 94]. This basically means that IPng has to provide services other than those already provided by IPv4 e.g. resource reservation, mobility, autoconfiguration and security. Obviously these services have particular market segments in mind. Commercial users would welcome the security [Bellovin, 94] and mobility aspects [Simpson, 94], whereas network entertainment users might welcome autoconfiguration and resource reservation [Droms, 93] [Taylor, 94].

The cost issue is a tricky one, as new IPng users will most likely see no immediate benefit over the use of IPv4 relating to this aspect. In fact, it may well cost them more as they will have to be trained in the use of IPng. There is, however, one major cost saving which can be incurred, that is the cost associated with network administration. Should IPng support remote resource management and better semi-automated network configuration (in particular routing and addressing setup), this will incur a large saving in the form of network administration costs. The current internetworking world requires that most of this configuration be performed by hand **using** text based editors and requiring a very good knowledge of the network topology and other important network properties such as size, protocol, physical medium etc. , making it extremely time consuming and hence very expensive [Gross *et al,* 92] [Dixon, 93].

The last motivating factor, namely the provision of connectivity to hosts which would otherwise be unreachable, depends to a large extent on the deployment of IPng and the state of the IPv4 address space in the near future. Most IPng proposals permit the interaction of IPv4 and IPng until the point where IPv4 addresses are depleted, after which stage some proposals al low for limited interaction between old and new protocols [Hinden, 94b][Hinden, 95][Curran, 94]. For

existing IPv4 sites there is currently little tangible motivation for them to consider migration to IPng as there should be continued connectivity till IPv4 address depletion, unless of course these sites wish to make use of the added functionality of IPng.

6.1.2 Backward compatibility

One has to consider why backward compatibility is an issue which has received so much attention by the IPng candidates. Consider the growth of the Internet (See chapter 2) and realise that there is a tremendously large user base which is already in existence. It would not make very good business sense to ignore this existing user base by introducing a new protocol which effectively isolated the existing IPv4 users. Any protocol which did this would have a very poor chance of surviving and would more than likely be nothing more substantial than a poorly supported and sparsely implemented curiosity [Claffy *et al,* 93].

In the evolution of microprocessors, in particular the Intel 80xx and 80x86 ranges, we have been taught the lesson of maintaining backward compatibility. These CPUs were manufactured with the specific intention of providing binary compatibility to their predecessors, and in so doing maintained their market share and user base [Hinden, 95].

Providing backward compatibility is going to have to be a primary design parameter of lPng. Backward compatibility can not be implemented as an afterthought, since it involves most facets of any new protocol. Consider the scenario where IPng could be a new packet switched protocol which has a fundamentally different packet size, header, address, routing protocols and offers
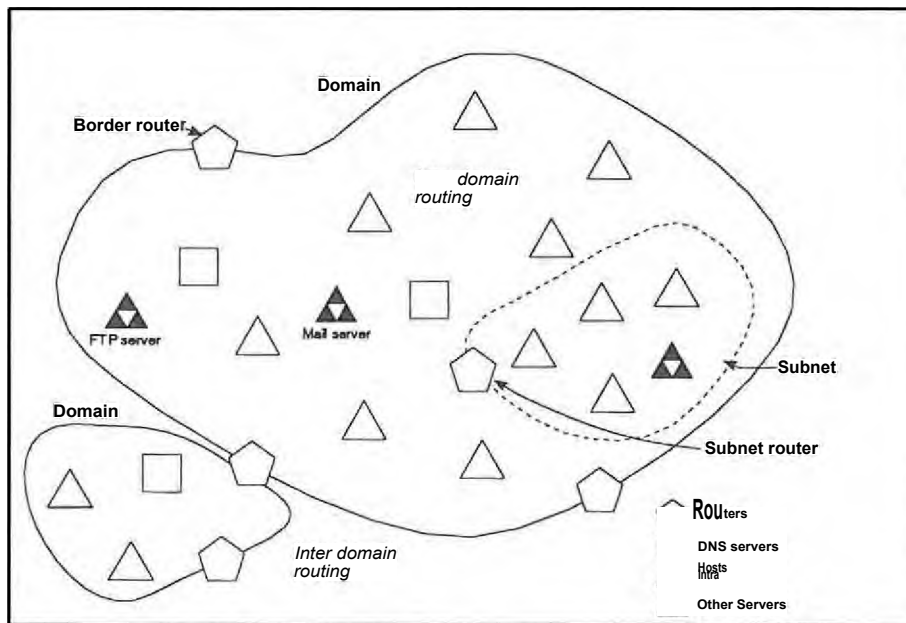
Figure 10 : Some of the fundamental elements of the Internet,

different services. The simplified internetwork as illustrated in **Figure 10** shows what the more fundamental elements in any internetwork are and hence how they will have to be modified to support IPv4 and IPng protocol inter-operability.

Providing for the most basic backward compatibility to IPv4 from this IPng would involve at the very least techniques [Hinders, 94a] such as:

- Extended DNS tables

- Address mapping/encapsulation

- **Header translation/rewriting**

- **Packet tunnelling**

- **Dual protocol stacks on routers as well as some hosts**

For the foreseeable future there are going to be IPv4 users, due to the fact that a large number of simple tasks or users do not require the added functionality of IPng, e.g. printer servers. For this reason, backward compatibility must not be considered to be a temporary state which IPng has to undergo but instead as a permanent feature which it must offer.

6.13 Incremental changeover

The Internet is a global entity and as such any attempted simultaneous protocol transition **has a very low chance of succeeding. Protocol migration would be a better term to use, as migration** implies a number of important aspects. The new protocol must be able to be deployed in **a diffuse fashion at the users pace.**

Using lessons learnt from the past, any changes should be implemented as a standard software upgrade, which can be performed with as little technical knowledge as possible. By providing the new protocol as a software **upgrade the transition is completed when the user decides it is necessary or required.**

An incremental transition also has the advantages that bugs in the transition process and protocol can be ironed out by the more technically inclined users who **will in most** cases **be the** first to utilise **IPng. In so doing they will be able to make refinements to the transition mechanisms and hopefully provide an easy and user friendly transition for the general users of the Internet.**

Unfortunately there are a number of inherent upgrade interdependencies in any network. In most cases the DNS servers, followed by the routers will have to be upgraded before the

users/subscribers will be able to upgrade their host machines. This places the limitation on the user that, before they can upgrade, the network administrator for their network has to have upgraded other crucial components in the network.

Problems that could arise with incremental deployment of a new protocol could be related to unforseen interdependencies beyond the immediate network domain, or hardware requirements being exceeded when trying to run some of the new protocol conversion schemes on some critical network devices.

It is difficult to foresee what other problems may arise relating to this incremental transition, as most of these problems will only become evident when the transition takes place.

6.2 Phase 2 : Why adopt a new protocol ?

The adoption of any new Internet technology will be predominantly market driven which means that it has to provide new capabilities or reduced cost to succeed and be accepted [Curran, 94]. In particular, the viability of a new Internet protocol has to be calculated by comparing the deployment difficulty as opposed to its potential benefits. This section discusses the benefits which a new Internet protocol should offer to the general Internet user.

6.2.1 Better support for new application types

There can be no doubt that the development of more diverse and complicated applications which run over the Internet will continue in the future. A successful Internet protocol is one which fosters the development of such applications, be they similar to previous applications or totally different. In an attempt to establish some uniformity and a common ground on which application developers can base their efforts, data exchange methods and a common data interchange format would be very useful [Clark *et al,* 94 This would facilitate the inter-operation of completely different applications with a minimum of effort or disruption.

la Common Data Interchange format

Data can be of a number of different formats. Typically in the past we have associated interchange formats with meaning things such as graphics formats or text formats. In more recent years audio formats and a number of hybrid formats have been added to these two. In order to make the growing amounts of information on the Internet more accessible it would be of great importance that the data which is resident within the Internet he compliant with one of the common interchange formats. A minimum list of formats which should be supported by any Internet protocol are:

+Text - The most stable and commonly used format, but will have to he carefully reconsidered as the need for character sets other than ASCII become appropriate due to the international flavour of the Internet.

+Image - Image representation

+Graphics - 2D Vector graphic information

+Video - General video format

+Audio - This should be for audio only applications and not a complementary format for video which already contains an audio track(s).

+Display - Typically display formats refer to the formats associated with windows which one would like to be able to open locally or remotely. An example would be requesting to open an X-Window on a remote machine.

+Data Objects - To facilitate inter-process communication, fundamental data objects such as integers, reals, strings and booleans, as well as their derivates, have to be well defined [Clark *et al, 911.*

VRML Virtual Reality Modelling Language is a file format which is currently under development to be used for all 3D graphics over the Internet. It will allow users to both view 3D worlds and interact with them [Pesce, 95].

♦ **HTML -** HyperText Meta Language, the format associated with WWW documents has been defined and is currently in its third revision. This is an example of the positive effect which a common data format can have on the distribution of information across the Internet. The growth in the number of web sites and web documents on the Internet can attest to this and is clearly visible in **Figure 11** [Network Wizards, 95c] [Pesce, 95].
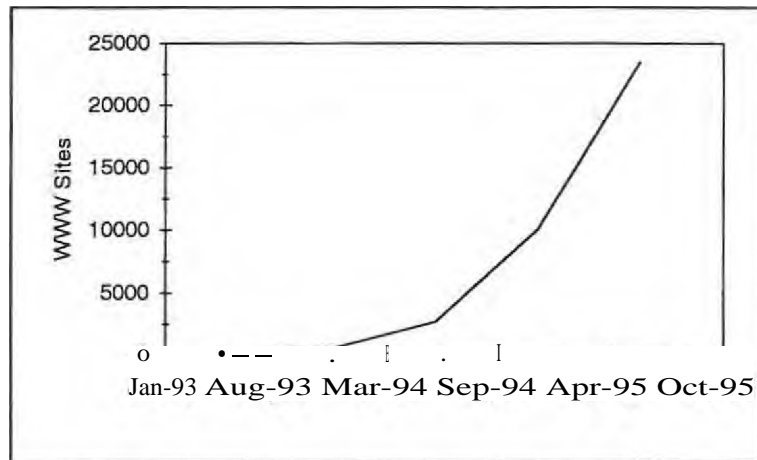
Figure 11 : Growth of the World Wide Web (WWW).

Application development should be guided by the definition of these interchange formats to **make the data accessible to** a wider audience than would be possible had each application required an in-house format.

❑ **Data exchange formats**

The existing Internet has been built on a packet based system which implements an

unreliable datagram service. On top of this the various protocols have supplied everything from unreliable delivery to virtual circuit protocols. The successor to IPv4 will have to provide a number of the same services as well as the ability to implement other services on the new **IP** protocol layer. A few of the required data exchange methods are listed below.

*Store and forward - This provides a typical system which can he utilised by hosts which have temporary network connection only.

*Global File Systems - Presently FTP is used to access documents at remote sites. Providing a global file service would mean that files could he accessed without using FTP. This would have the effect that the information in the files would be far more useful. For instance, operations which are currently possible over File Systems such as NFS would then to possible over the entire. Internet. (Not only read only access but also write access, execute access etc.)

*Inter-Process Communications - The field of distributed computing allows the exchange of data over a network. Support for this exchange of data over the Internet would mean that Internet resources would be able to participate in distributed computing without having to resort to a special protocol on an isolated network.

*Data Broadcast - Applications which need to receive broadcast updates for their continued operation would make use of a data broadcast. This method would have to be efficient and reliable.

*Database Access - A global file system could provide one with the data contained within a file, but it does not allow one to manipulate the tile structure or its database attributes, which a database access service would do [Clark *et al,* 91].

6.2.2 Advanced network services

Any new protocol will have to offer a minimum set of new and efficient network services which are currently implemented under IPv4. Most of these services have arisen due to a particular need, but due to limitations of the IPv4 protocol their implementation and operation leaves a lot to be desired. Services such as mobility support, data security, Type Of Service (T.O.S) considerations and autoconfiguration can be considered as the absolute minimum set which have to be provided [Gross *et al,* 92]. Each of these have their own specific requirements which will be explained briefly.

Mobility

Mobility places the requirement on the network of seamless operation as the host is in transit between networks, which implies some dynamic routing with as little packet retransmission as possible [Kleinrock, 94] [Droms, 93]. Host mobility also gives rise to interesting questions relating to addressing [Ramanathan, 95b]. Under the current IP protocol an IP number belongs to a particular machine. When **that** machine is issued an IP number, the number also contains locational information in the form of the network part of the IP number. Now as the machine/host moves around within the Internet the network part is no longer valid and hence some specialised routing has to be performed. Using IPv4 and catering for mobile hosts puts tremendous strain on routing resources. This is because they are normally catered for by having a single entry in the routing tables for each mobile host, effectively punching a hole in address aggregation schemes.

Security

In native IPv4 there is very little provision made for the transmission of secure datagrams. The need for confidential and secure datagram transmission is emphasised when considering the commercial nature of the Internet and its current growth trends [Curran, 94]. There are applications which provide some sort of end to end security. However, it

would be very useful if there was a set of standard encryption techniques w hich could be applied at each of the protocol levels providing the different levels of security. These security techniques should be simple enough that they can be properly understood, yet complex enough that they are sufficient to protect against attacks. The protocol layers **at** which security could be implemented are the link layer, net layer, host level, and application level. The requirements and assumptions relating to security change as one moves down this protocol stack, generally increasing the size of the security perimeter[5] **[Clark** et al, **91].**

❑ **Type of Service (T.O.S.)**

There are a large number of applications which have special requirements relating to **latency and throughput for their operation. With IPv4 they have to** be happy with the **performance which they get, irrespective of what their needs** might be. The idea of allowing some sort of underlying resource reservation to permit a guarantee of a **certain level of service provision has been suggested [Clark** et al, 91] [Gross *et al,* 93]. This would then permit applications to run with better performance and generally more reliability. Consider, for instance, the transmission of video during a real -ti me video conferencing session. Here the level of service provision would have to guarantee **enough bandwidth for the video transmission as well as a** low latency so as to prevent glitches in video display.

❑ **Autoconfiguration**

Autoconfiguration can be coupled with the need to provide support for mobile hosts. However, there is also a need for autoconfiguration in the more conventional markets where the users are not likely to be technically inclined and want to be able to get the use of the Internet without having to first become familiar with its technical side. This network

---

[5] Security perimeter - the components of a network which are secure. Typically the hosts **a domain could be** a security perimeter.

service is vital to the long term growth of the Internet, as without it there is going to be a very severe limitation on who can connect to the Internet and who cannot [Dixon, 93] [Curran, 94].

6.2.3 Continued long-term connectivity

The transition to a new protocol can guarantee long term connectivity, which is not the case for those hosts which decide to remain IPv4 compliant only. Although some degree of backward compatibility and inter-operation is planned, there is likely to be a point at which only a very small subset of the existing IPv4 address space will have connectivity to the rest of the lPng Internet. This is related to the final depletion of the IPv4 address space and how this space is used to provide backwards compatibility.

A simple scenario would be as follows. For an IPng host to be IPv4 compliant it needs a temporary IPv4 address. It can only get such an address if the IPv4 address space has not been exhausted. By definition, after address space exhaustion there can be no inter-operation unless the IPng host is one of the few privileged hosts which has been assigned both a permanent IPv4 address and IPng address. No doubt there will be ways to circumvent this problem by performing translations or using a protocol conversion gateway, but it is very unlikely that the full functionality will be maintained during any of these processes.

6.3 Phase 3 : The underlying advantages of IPng.

6.3.1 Address pool size

Any new IPng protocol has to make allowances for future Internet growth, and so doing provide for a considerably larger address pool. This address pool should have some underlying structure for the distribution of addresses which can then be used to make routing decisions easier. These underlying structures or aggregations can be along administrative domains (AD), physical

locations, or any other logical aggregation characteristic.

Consideration will have to be given to the various issues surrounding IP addresses, in particular those related to renumbering strategies, service provider numbering and aggregation policies.

Renumbering strategies have been proposed by a number of protocols [Chiappa, 94a] [Deering *et al,* 94c] [Gerich, 93]. These proposals have come under severe criticism, especially relating to the large amount of effort required to renumber an internetwork the size of the Internet. Despite this criticism, renumbering is seen as one of the possible solutions to some of the growth problems within the Internet. In particular, it is considered to be a viable solution in areas where the original provision of well aggregated addresses' has been outgrown. The idea is to renumber all the hosts within this area, separating them into a number of new and larger aggregated groups along some new criteria. This would have the effect that routing tables for this domain would only have to advertise the appropriate address prefixes for each aggregation and not be overloaded by having to advertise a large number of unique addresses which did not fall into the existing addresses aggregations [Gerich, 93].

Service provider numbering is another area which has come under criticism. Strong opinion has been voiced as to whether service providers should get a chunk of the new address space and then be able to assign addresses to their clients from this space [Casteneya *et al, 94] [Casteneya a et al,* 95]. Despite it being a good aggregation criterion it does have the nasty effect that should a user decide to change service provider, there would probably be a required address change as well. This might be seen as a reasonable requirement by some, but it must be remembered that in any sizeable organisation it would be an enormous task to perform un-aided address renumbering for all the hosts within that organisation. In effect, if an easy method for address renumbering   were provided as mentioned earlier, then the enormity of this problem would diminish.

---

[0] Aggregated addresses - typically address which have some common characteristic, for instance they may have the same network number or a certain number of high order bits which are the same. Aggregation in this sense implies grouping based on a shared characteristic.

Aggregation policy is another area which will have to receive careful attention. Essentially, the reason for performing address aggregation is to reduce the load on the routers, which have to perform routing for the various domains within the Internet [Claffy *et al,* 93]. Even if an IPng candidate reduces the load on routers considerably it would still be irresponsible not to perform some type of address aggregation. As mentioned before, this aggregation could be performed along a number of lines, such as administrative domains (ADs), service providers, governments, physical location, continents or a number of other characteristics. A key issue to the efficient operation of the Internet will be maintaining these aggregated groups and ensuring that there are a minimum number of hosts which have special un-aggregated addresses advertised in any routing tables [Clark *et al,* 91].

6.3.2 Network Pool size

Coupled to the growth in the number of network hosts is going to be the growth iii the number of networks. At this point we consider networks to be related to the physical entities which are also referred to as networks. It is not unlikely that in the future, homes which are linked to the Internet, making use of Internet device control and entertainment as well as the more "standard" education and computer access, could be seen as complete networks. With this in mind, it has been estimated that any IPng must support a growth in networks up to at least 10^9 unique networks [Clark *et at,* 91].

This is a very large growth which immediately makes apparent the fact that the Internet is going to require routing protocols which are far more efficient that those currently used. Currently the core Internet routers are handling tens of thousands of networks. These numbers are already placing incredible strains on the routing resources, both its software and its hardware. Ideally, we would require some sort of hierarchy within the network numbers as well as the address aggregation to solve some of the tremendous overhead associated with routing [Gerich, 93] [Clark *et al,* 91].

6.3.3 Routing hierarchy

Providing an efficient and effective routing hierarchy for the Internet has to he one of IPng's primary objectives, as this is currently the most serious problem with IPv4. The two tiered hierarchy as provided by the existing 32 bit IP address, which also functions as a selector', is not sufficient for the number of networks in the Internet. With a two level hierarchy the number of networks which have to be advertised at the higher hierarchy level in core routers is too large for these routers to handle. it is not simply a problem of inadequate hardware, but rather one of unforseen growth in networks and hence an inability to cope with the current scale. The hardware being used is state of the art, which indicates that this problem can not be solved by throwing more hardware at it. Instead, a fundamental change in the routing hierarchy structure is required [Claffy *et al,* 94].

lPng will have to provide a hierarchy of some sort which has more levels within it, so as to relieve the burden on the core Internet routers and spread the routing load more evenly across all routers. A number of proposals with regard to routing strategies have been made.

There have been suggestions made that the routing should be done in two regions within any one domain, with border routers which have small but very fast caches and which send a route request to some centralised router server if they don't know the destination. This design stems from the fact that routing as we know it can be separated into two distinct operations, the fi rst being route calculation, and the second being the forwarding of datagrams. Since simple forw a rdi ng is an inexpensive and hence a fast process, this operation has been assigned to the border routers. The route calculation has been assigned to the dedicated router server. This scheme has the advantage that for a domain which uses a small set of routes, the routing will be very fast, as most of these well-used routes can be stored within the small caches of the border routers. One problem which is soon apparent with this approach, is that using this system there is a single point of failure,

---

[7] Selector - information which is used by routing devices to determine how to handle the packet. ln TCP/IP some of the high order bits in the 32 bit I? address are used to determine the route for the datagram.

namely the router server. This need not be the case as a route server could he replicated across a number of different machines within the domain, making the system less prone to failure.

Another suggestion is to use a scheme of distributed routing across the entire Internet. The approach is to have routers query their neighbours if they do not have the route in their own routing tables. Obviously this approach means that there has to be an efficient mechanism for the querying of neighbouring routers and a robust algorithm for the distribution of rout i information.

A fundamental issue regarding routing is to make some sort of network aggregates' the units **for** routing. The temptation to provide a temporary solution by using individual hosts as routing units must be avoided at all costs.

6.3.4 Network and resource management

Network management for the existing TCP/IP Internet is very time consuming due to the large amount of time which has to be spent performing a relatively small variety of tasks. usi ng manual tools. Typically, the issuing of addresses within a network as well as router configuration are very time consuming issues and require a great deal of a network manager's time [Gross *et al,* 92].

IPng should provide powerful network management tools or, at the very least, hooks within **the** protocol so that these tools can be developed. A major factor influencing the accepttance of IPng is the ease of management by the network administrator. They are, after all, the people who have to concern themselves with the day to day operation of the new protocol [Clark *et al, 91*]. Included in these tools must be the capabilities to perform remote configuration of devices from some central point with the domain. This is especially useful in the setting up of routers in a network. Monitoring software for network traffic is also a priority [Saunders, 95]. A further

---

8 Network aggregates - A group of networks which have some common characteristic and .   located in the same administrative domain (AD).

requirement for network management under IPng is to provide tools with which network statistics can be gathered. This is important due to the fact that network optimisation can only be performed properly if one has a knowledge of what is happening on the network.

Any designs for network management tools must also bear in mind that the Internet is going to remain, for the foreseeable future, a diverse mix of commercial, government, academic, educational, and research networks, each with their own particular characteristics. By their very nature, networks used for different purposes have different requirements . The tools created for managing these networks must be flexible enough to be able to satisfy the needs of any one of these networks.

6.3.5 Accountability

With the trend in diversification from purely academic and research networks into the environment of commercial networks [Claffy *et al,* 93], a higher emphasis is being placed on being able to record resource usage. These recorded statistics can then be used to charge clients for their access etc. To be able to perform this recording, one has to consider what the resources are which should be monitored. Typically, monitoring of server access and login time is one, another perhaps less obvious one is monitoring the traffic sent through any particular router by clients. To monitor router usage implies that some additional state information is going to be stored in the router. Typically, the router will have to keep some record of which flows are passing through it, their sources and utilisation details [Clark *et al,* 91]. A rather unfortunate side-effect of this monitoring is the extra processor time which will be used doing nothing constructive for actual datagram routing.

6.3.6 Data integrity and security

Despite data security being discussed in section 6.2, it also has important implications for the system administrators. Also mentioned in section 6.2 is the fact that there are a number of different protocol layers at which this data encryption can be placed, each providing some sort of

security perimeter [Clark *et al,* 91]. For system administrators data integrity is important, especially when considering network operations such as remote login, system kernel integrity, user validation techniques and firewalls. Data integrity i: especially important when considering the threat which computer viruses can have on computers in general. Likewise, data integrity and security can be threatened by crackers', who find a challenge in breaching security in systems, often creating havoc in the process.

6.3.7 Multi-Protocol architecture

The Internet is and will remain a multi-protocol architecture. This means that IPng will have **to be** aware and tolerant of other protocols running over the same physical network. Tolerance also implies minimal interference and ideally maximum inter-operation, should it he requested. The handling of shared system resources is a point which has to be considered here. For example, consider the case of a dual protocol Internet, where the one protocol (Protocol A) implements resource reservation and the other (Protocol B) does not. Protocol A should he aware of Protocol B and ensure that at no time has it performed resource reservation to the extent that Protocol **B** is unable to operate successfully [Clark *et al,* 91].

6.4 Conclusion

No internetworking vendor can afford to deploy anything in the marketplace which is not desirable by consumers. In order to be successful, IPng has to offer clearly improved functionality over IPv4 and at the same time offer some sort of transparent access between I Pv4   Png once 32 bit address depletion occurs. As mentioned throughout this chapter, this added functionality must be offered not only to the general Internet user, but also to the technical user a nd perhaps most importantly to the network administrators.

Each of the proposed candidates for IPng will be evaluated against the criteria set out in this

---

[9]Crackers̶ people who intentionally break into computer systems illegally, often with   intent.

chapter and the relative merits and shortcomings noted, with particular attention **being paid to** the routing and addressing related issues. It is very unlikely that any one **protocol can solve all** the problems, it is more likely to be a case of constant trade-offs, to reach a cony' **Soluti**on.

7. Operation of the Proposed IP Protocols

This chapter contains a number of the proposed protocols which address some of the existing problems with IPv4. Each protocol is examined briefly and the implications that it has on routing and addressing are explored. As the subject of this document is related primarily o addressing and routing issues, these are the two aspects which are given the most attention when examining the protocols.

The protocols presented in this chapter have been proposed over several years. typically the last four years. During this time it has been evident that there have been several changes in the requirements for a new **IP** protocol. However, the requirements concerning routing and addressing have not changed much at all. There has been and still is agreement that the addressing issue requires more space for future expansion, and that this space has to he organised to place a minimal burden on the routing infrastructure.

Several of the protocols do not attempt to solve both problems, but rather specialise in solving one or the other well and sometimes leave the appropriate hooks so as to al low combination with another protocol to solve the remaining problem.

After an overall description of the protocol, the author presents a cost and benefit analysis for each protocol, and indicates how the particular protocol matches the criteria set out in Chapter 6, particularly Section 3. An indication is also given as to how the protocol approaches solving the problems which are present in Chapter 3.

7.1 Classless Inter Domain Routing (CIDR)

CDR is primarily interested in the large scale IP address allocation within the Internet and deals with the following issues:

• Benefit of encoding topological information into an IP address to reduce routing overhead.

• Anticipated need for additional hierarchy within the Internet to support network growth.

• Recommended mapping between Internet topological entities and IP addressing and routing components_

• Division of **IP** address assignment among service providers.

• Choice of a high order portion of IP addresses in leaf domains' with multiple service providers" [Rekhter *et al.,93].*

CIDR proposes a form of topological IP address assignment that can he used for effective address abstraction within the Internet. CIDR further suggests that there be distributed address allocation along some guidelines that will provide a mechanism for the aggregation[12] of routing information. Furthermore it is suggested that hierarchical routing be used to achieve some sort of routing data abstraction[13] at the inter-domain levels in the routing hierarchy. The abstraction of reachability information[14] dictates that IP addresses should be assigned according to topological criteria. The

---

[10]Leaf domains - sites which are the end users of a specitic domain.

Service providers - domains which share their resources with other domains.

[12]Aggregation - grouping of information according to some common property or attribute network numbers by their provider prefix or high order bits.

[13]Routing data abstraction - summarisation of routing data which in turn allows the use or less resources during the routing process_

[14]Reachability information - describes a set of reachable destinations

nature of IP address assignment is such that it has in the past unfortunately often been done by political or organisational boundaries instead [Gerich, 93].

If IP addresses within some domain" are drawn from a noncontiguous IP address space, then the information exchanged across the routing boundary consists of an enumerated list of IP addresses. Alternatively, should the IP addresses within some domain be drawn from a contiguous IP address space, they can be adequately advertised as a single IP address prefix. To take the analogy one step further, if routing domains were organised in some form of hierarchy, the advertising of routing information could be abstracted as it propagated up each level in the hierarchy.

In CIDR, addresses are quoted as <Address prefix : CIDR mask> tuples where this format determines the base address for an address space and also the size of the address space, for further illustration of how this works see **Figure 12.**

Interpreting a CI DR Tuple

<Address Prefix : CIDR Mask >
   096.128.1.0 : 255255.248.0>

Base address   196.128.1.0

                      255.255.255.255
                    - 255.255.248.0
Size of address block   7.255

Valid addresses   196.128.1.0
                      196.128.1.1
                            .
                      196.128.8,254
                      196.128.8.255

Figure 12 : CIDR Address Tuple

In **Figure 13** an example is given where **IP** addresses are allocated from within a provider's address space. The addresses are then used as convenient to the subscriber, either all as host addresses as in **Figure 14** or split into a further two subnets and their associated   host addresses in

---

[15] The resources under the control of a single administration

Figure 15. The important thing to notice is that in both cases only one **<address prefix CIDR mask>** tuple has to be advertised for the provider. CIDR differs from conventional subnetting in that it is classless and can allocate address space across conventional class boundaries, without the additional overhead of having to advertise each network number for each address class used.

BACKBONE

U

<198.128.0.0: 255255.0.0>   **Provider** Prefix

U

<196.128.1.0 **: 255255.248.0>**   Subscriber Prefix

Hosts in subscriber domain   Subnets in subscriber domain
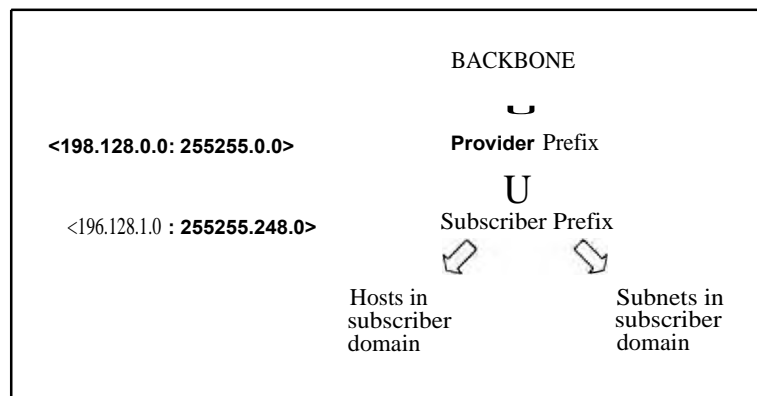
Figure 13 : Prefix abstraction using CIDR(most specified prefix is at the•
level and the least specified prefix is at the provider level.)

**Subscriber
<196.128.1.0 : 255.255.248.0>**

**1961 28.1 .0
196128.1.1**

**196.128.8.254
196.128.8255**

Host addresses
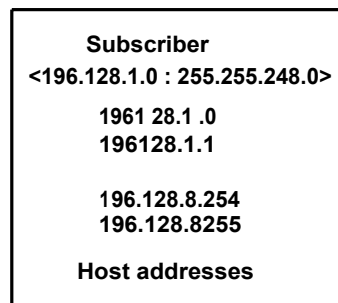
Figure 14 : Using the complete
subscriber address space for hosts
without subnetting.

Where little or no abstraction can occur, a flat routing space results for int er-dk imam 'outing (the current situation with TCP/IP). This flat routing space requires that all routing domains have explicit knowledge of all other domains to be able to route traffic to them [1':iste!, 81].

```
┌─────────────────────────────────────────────────────────────┐
│                                                               │
│         Subnet 1                        Subnet 2              │
│  <196.128.1.0 : 255.255.252.0>   <196128.5.0 : 255.255252.0   │
│                                                               │
│         196.128.1.0                     196.128.5.0           │
│         196.128.1.1                     196.128.5.1           │
│                                                               │
│         196.128.4.254                   196.128.8.254         │
│         196.128.4.255                   196.128.8.255         │
│                                                               │
│         Host addresses                  Host addresses        │
│                                                               │
└─────────────────────────────────────────────────────────────┘
```
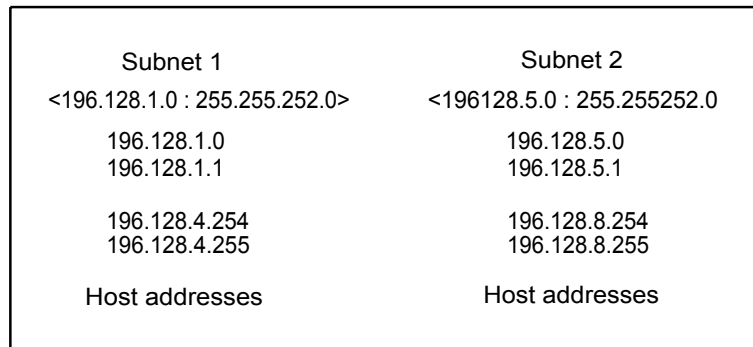
Figure 15 : Splitting the subscriber address space into two subnels have the equivalent size of 4 class C networks.

7.1.1 What does CIDR deal with?

CIDR deals directly with the problem of IP address allocation and IP packet routing costs [Rekhter *et* a/ ,93a]. IP addresses are divided into address classes, the most common of these being class B and class C [Postel, 81]. As was seen in chapter 3, these classes have associated with them an upper bound as to the number of hosts that they can support. Unfortunately a single class C address often does not provide enough address space for a medium sized network. Therefore, these organisations try to get a class B address allocated to them. Class B addresses are however scarcer than class Cs and therefore difficult to get assigned. As :1 solution to this problem, multiple class C addresses are often assigned to an organisation 111.1i is going to need more than 2^8 hosts on its network. This in turn means that an address prefix network number) has to be advertised for each class C address in the routing tables. Multiple address prefix advertising for a single organisation or network is wasteful and uses valualile resources in routers. Another side effect of this address allocation mechanism is that there is norm;:ily a large proportion of the allocated address space that is never used, especially with CI:Ns B addresses.

Deploying CIDR can be used to slow the growth in routing tables in coin pm 1, in to the number of allocated networks. It also has the advantage that it does not have to be clyployed globally to function, but can be deployed incrementally, while still providing full conncciivity.

7.1.2 How does CIDR work?

As noted in the introduction, CIDR has an implementation plan that affects three distinct areas of interest, namely;

• mechanism for aggregation of routing information

• distributing the allocation of IP address space

• address space aggregation efficiency

These three areas are those of relevance to this work and will now be den!! ' lilt in more detail.

CI Aggregation mechanisms and their shortfalls

Before we can consider aggregation mechanisms we have to consider the different types of organisations that we are likely to find in the Internet, as each of these is likely to require an appropriate aggregation technique. Two common types 01 oi ganisations are single-homed[16] and multi-homed" organisations [Fuller *et al* ,93].

Single homed organisations which use a subset of their Ti:i nsit Routing

---

[16] Single-homed organisation - Has a single **TRD** connecting its network(s) to the I 011,-1 lick

[17] Multi-homed organisation - Has multiple TRDs connecting their networks to thi   rnet at Jae$_r$ent physical locations.

Domain's' (TRD) address space pose no problem for agw ctil on as they can aggregate their addresses into a single prefix which forms n sunset or their TRD's address space. This principle is illustrated in **Figure 16.**
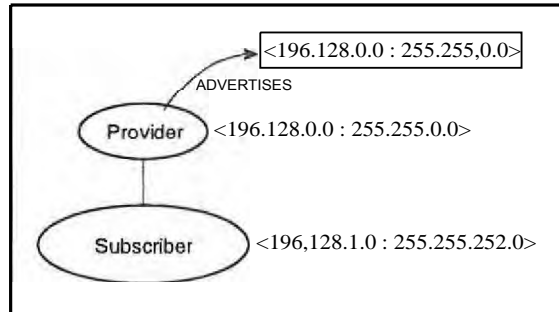
Figure 16 : Single-homed organisations using CIDR

Multi-homed organisations are more complicated as they have multiple attachment points with different addresses from different TRDs. See Figure **17.** Multi-homed organisations' address prefixes need to be explHily advertised by each connected **TRD,** because of their multiple attachment points. In this case both Provider A and Provider **B** have to advertise the address tupic 196.128.1.0 : 255.255.252.0>, even though it falls within Provider A's attacgnie. Although this may seem a waste of valuable resources, the number of multi-It imed organisations is in fact very small, and unlikely to grow considerably in tilt. suture [Rekhter *et al,* 93b] .
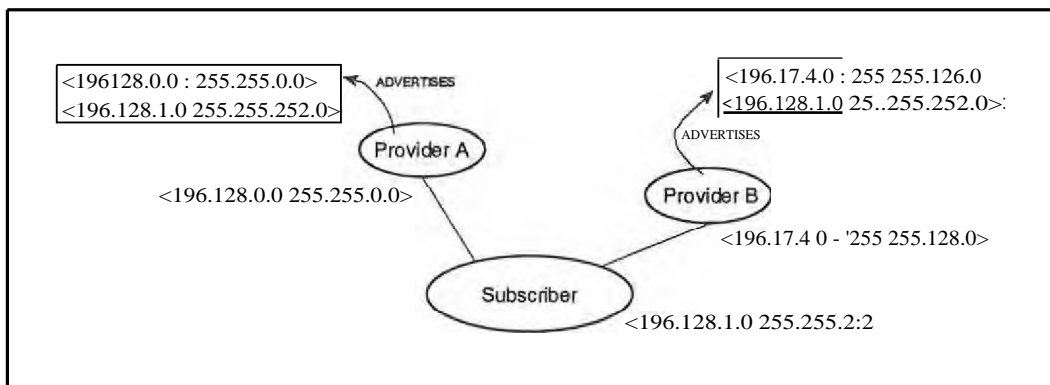
Figure 17 : Multi-homed organisations using CIDR.

[tar] transit Routing Domain (TRD) - A domain which provides connectivity between ;1 .1..   :11)er ǀ twurk and other networks in the Internet. Also known as a service provider.

For multi-homed sites, aggregation may occur at some higher level i it lite network hierarchy. For instance if two sites belonging to the same organisation are connected to the NSFNet, both obtain their address space from NSFNet and as a rcsi ilt can be aggregated at the **TRD** level.

Organisations that change TRDs but do not renumber pose a problem Hr address aggregation. These organisations effectively break the aggregation which existed. As this scenario is likely to be a common one, there is a strong motivation Hi' dynamic host address assignment to ease the process of migration from old to new ;addresses [Droms, 93].

There will always be some degree of aggregation possible for sites that are allocated **a** contiguous power-of-two block of network numbers, as their addresses can all he represented by a single prefix and mask pair [Rekhter *et al ,* 93a].

ZI Distributed allocation of address space

CIDR requires that contiguous blocks of class C addresses are allocated to service providers by the Internet address issuing authority (InterNIC). The *service* providers in turn allocate bitmask subsets of these to their subscribers'. Because ()I' the deployment of classless network inter-domain protocols like **BGP-4** and IDRP, this hi kvise masking can now be used for allocating address subsets instead of address classes [Fuller *et al* ,93]. Subscribers will have their address space allocated from within their provider's address space as shown in **Figure 18.**

CIDR can use the entire 32 bit IP address space for supernetting' of class C or

---

[19] Network service subscribers - Domains which utilise other domains resources,

[20] Supemetting - allocating blocks of addresses from a contiguous block of addressc It; .itay si);111 no re than one conventional IP address class. Typically Class C addresses *are* used for superrie

subnetting[11] of class A or B [Fuller *et al,* 92a]. Some changes are required to various aspects of the routing protocols before Class A addresses can be subnetted effectively, these will be discussed later in this section.

BACKBONE

Provider
<196.128.0.0 : 255,255.0.0>

Subscriber Prefix
/ <196.128.1.0 : 255.255.248,0>

Hosts in subscriber domain
<196.128.1.0>
<196.128.1.1>

<196.128.8,254>
<196.128.8.255>

,Subscriber's address space r'
Provider's address space

=Part of the larger domain
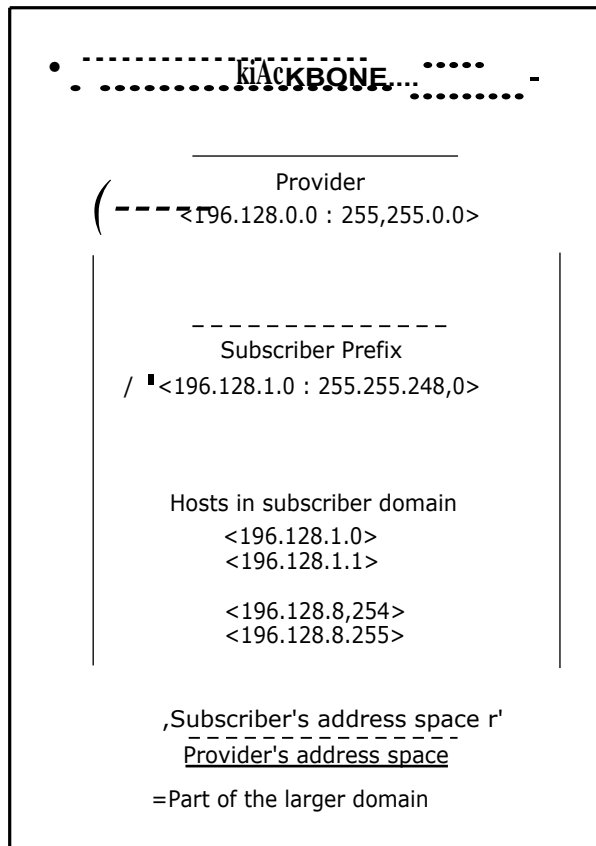
Figure 18 : The relationship between provider address space and subscriber address space using CIDR.

Having address space allocated by InterNIC to service providers and not directly to the end user is easier and provides a distributed address allocation scheme. It also forms the basis of a good delegation method for the future allocation of network numbers.

---

[21] Subnctting - allocating multiple networks address blocks from within a conventional EP address class. Typically class A and B addresses are viable address classes for subnetting.

❑ **Address space aggregation efficiency**

The level at which address aggregation is performed in the hierarchy has direct bearing on the gains one can expect as one progresses up the hierarchy. This relation between the gain and level is almost exponential as is shown in **Figure 19.** What is required is to find the balance between efficient routing and decentralised address administration.
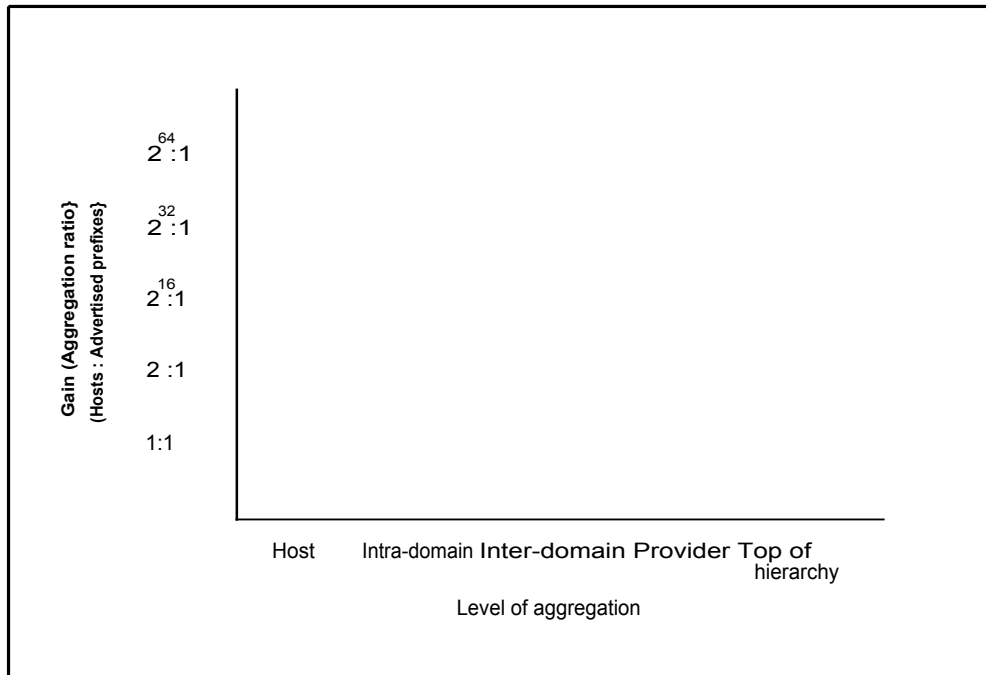
Figure 19 : The benefits of address aggregation and how they depend on where it is performed in the network hierarchy.

The major components of the Internet are service providers (e.g. backbones and regional networks) and service subscribers (e.g. sites and campus networks), which are arranged naturally in some form of hierarchy.

There exists a natural mapping from these components to the IP routing components,
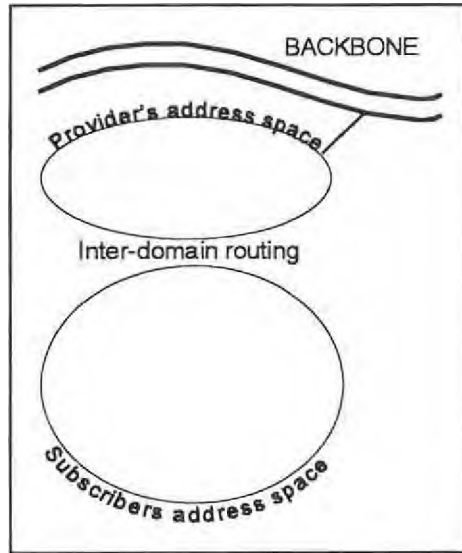
Figure 20 : Inter-domain routing due to subscriber using a separate address space.
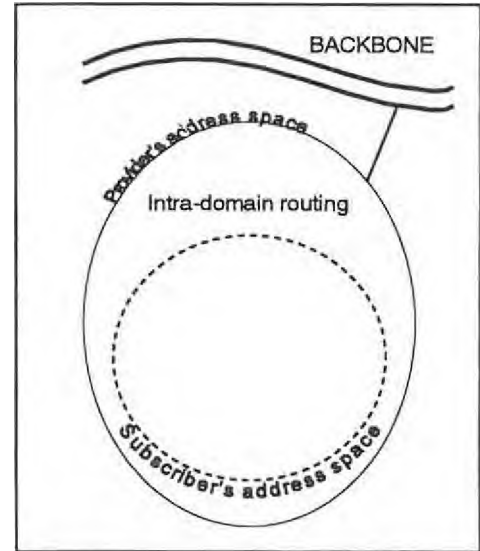
Figure 21 : Intra-domain routing with the subscriber using a subset of the provider's address space.

where subscribers can either act as separate routing domains or share the routing domain of their provider. See **Figure 20 and Figure 21.**

Considering the major components as mentioned above one has to decide where best the data abstraction and administrative decentralisation should be performed. The options are:

- within the routing domain

    (i.e. subnetwork or router)

- in the leaf routing domain

    (i.e. at the sites)

    Leaf routing domains are concerned with the intra-domain routing services.

- in the transit routing domain

    (i.e. backbones and providers that are both forms of TRDs)

    TRDs are concerned mainly with the carrying of transit traffic.

PAGE 63

• at the continental boundaries [Rekhter *et al ,93]*

Other factors which also have to be considered are that the greatest burden for transmitting and processing routing information is at the top of the routing hierarchy. Also, the advantage incurred because of abstracting is small to the layer performing it, since it still has to maintain information for all the attached topological routing structures.

Looking at the above four alternatives in more detail enables one to determine where best to place the abstraction and administration mechanisms.

+Administration of IP addresses within the domain

This is exactly the case at the moment, where there is no common prefix and all
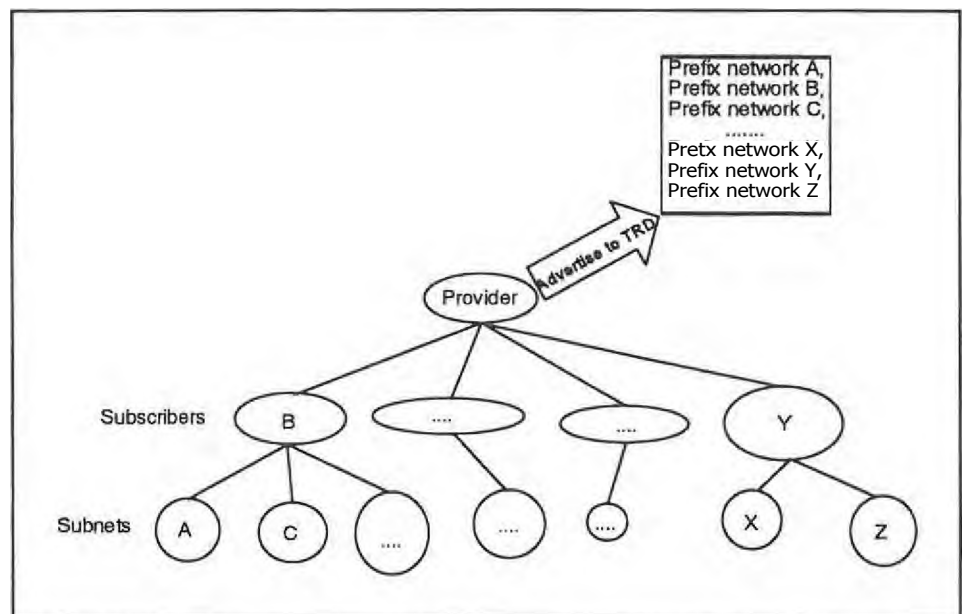
Figure 22 : Flat routing and individual network number advertising.

the subnetworks have unrelated IP address space assigned to them. As a result there is a flat routing space, [Postel, 81][Rekhter *et al,* 93] which is still routed according to address classes. The leaf domains have to advertise an order of the number of networks attached to them. This then propagates upwards and the

provider has to advertise all the network numbers attached to all its subscribers' leaf domains. Finally, at the backbone level all the network numbers for all providers have to be advertised. Clearly this is not a good solution since the routing tables at the backbone level will have to be extremely large to cater for all these entries. This case of flat routing, as described, can be seen illustrated in **Figure 22.**

*Administration of **IP** addresses at the leaf routing domain
This solution provides the greatest level of abstraction possible as this is the lowest level in the hierarchy. Each leaf domain (site) could obtain an address prefix from its provider and then only advertise this single prefix to the provider. The provider would in turn only advertise a single prefix for each of the leaf networks.

*Administration at the Transit Routing Domain
There are two kinds of TRDs, namely direct and indirect TRDs. Direct TRDS act solely as service providers whereas indirect TRDs have subscribers who themselves act as service providers. Indirect TRDs are also known as backbones. Multi-homed routing domains have to be considered at this level as they place certain routing requirements on TRDs as were shown earlier. Each of these cases will now be briefly outlined and it will be illustrated how they each affect address aggregation in the network hierarchy.

*Administration by Direct Service providers
The direct service providers would issue part of their address space to their leaf domains. This means that they would advertise a single prefix for all the leaf domains connected to them. Consider the example shown in **Figure 23.**

e.g. Consider the case of a]eaf domain with SANet as the service provider. SANet has IP address/mask tuple of <127.1.0.0 : 255.255.0.0>.

 SANet could then allocate the leaf domain the IP address/mask tuple of <127.1.0.0 : 255.255.240.0>

This would give the leaf domain an address space equivalent in size to 16 class C addresses (i.e. 16x254 **IP** addresses), with 11' addresses ranging from <127.1.0.0> to <127.1.15.255>.

Figure 23 : Example of a subscriber address prefix and mask tuple.

*Indirect providers

There is little benefit in direct subscribers taking their **IP** address space from within that of indirect providers' address space. This is substantiated by considering the limited number of indirect providers, and the likelihood of this staying the case. Furthermore, backbones are perceived as being independent which is another reason that the suballocation of address space from them is not advised. Direct connection by subscribers to the indirect providers is likely to increase to the extent where the indirect and direct classification becomes blurred [Rekhter *et al* ,93].

♦ Multi-homed routing domains (MHRDs)

Multi-homed routing domains (or domains attached to more than one service provider) have several possible solutions concerning address allocation, each of which will be examined in more detail.

1) MHRDs can have independently allocated (from TRDs) **IP** address space and thereby require explicit prefix advertising in the TRDs' routing tables. This solution results in packets sourced from outside the domain entering the domain closest to the source. This in turn maximises the load on the internal networks.

2) Separate IP address space can be allocated for each connection to the TRDs, based on the closest attachment point. This means that each TRD announces which parts of the organisation can be reached through its own address space. An important implication of this is that no additional routing information is required and therefore less information is needed in routing tables, which means that better routing table scaling results. This allocation mechanism results in packets entering the domain closer to the destination, maximising the load of the TRD. Furthermore it also allows for greater specification of policy routing, which cannot be achieved in the first case.

A serious problem can occur with this scheme, namely: If one of the TRDs goes down, traffic is not default routed intra-domain and is lost, unless explicit backup routes are defined in the routing tables. Furthermore, if there are any external changes (i.e. to the TRD) then the internal numbering also has to change [Rekhter *et al,* 93b][Fuller *et al,* 93].

3) Assign a single address prefix for a multi-homed organisation based on a single TRD connection. The other connected TRDs then maintain an explicit routing table entry for this connection and perform selective advertisements of these routes to other TRDs. This selective advertisement establishes a default route that all TRDs will know how to reach (since all TRDs are linked to each other).

+Continental aggregation

Continents provide natural boundaries, these being topological and administrative boundaries. Each continent can be allocated a subset of the global IP address space, which can be further distributed on that continent. Address aggregation

should be performed on the far side of the continent, to prevent excessive use of the transcontinental lines for error messages. (E.g. unreachable destinations get thrown out before they leave the continent of origin.) [Rekhter *et al,* 93a]

7.1.3 Cost and Benefit analysis of C1DR

CIDR was evaluated against the criteria listed in chapter 6, with particular attention being paid to the addressing and routing issues of section 3. Furthermore, some important issues relating to protocol life-span and protocol implementation costs were isolated. The evaluations of the above-mentioned aspects will be presented in this section.

CIDR will coexist peacefully with other protocols, to allow for a multi-protocol environment. CIDR was designed as a short term solution with an expected life span of 3 years [Fuller *et al,* 93]. Due to this design parameter a number of factors are not addressed by CIDR, for instance, policy based routing and other more advanced routing requirements. On the other hand being a short term solution required easy implementation. Consequently CIDR requires no substantial change to the topology or even the address assignment policies. As mentioned earlier, the renumbering of domains which do not have topological assignment will result in considerable benefits when using CIDR.

An approximate calculation of the cost in terms of routes which have to be advertised which occurs as a result of not using address aggregation schemes.

Providers	=2000	[Network Wizards, 95b]
Domains	z 120000	[Network Wizards, 95b]

— 60 domains per provider (aggregate tot address prefix 4- multi horned organisation preruces)   (1)

Hosts   ll6642000   [Network Wizards, 951e]

— 55 hosts per domain (aggregate to I address pea)

Routing tables at the various levels in the hierarchy may contain the following number of network addresses. Each explicitly advertised network number has a cost associated.

Unaggregated networks advertising at provider level : **60 routes**   (^{1})

Unaggregated networks advertising at backbone level   **n** * 60 (o.numberof providers connected to backbone)

Unaggregated networks advertising at core router level : n * 60 sm (m=ntunber of backbones connected to core routers)

Figure 24 : The cost associated with not using address aggregation within a hierarchical network.

Stemming from the analysis of this protocol against some of the criteria identified in Chapter 6, the author believes that before the full benefit of CIDR can be seen, there has to be a topological address allocation scheme and the deployment of CIDR capable inter-domain routing protocols. See the calculations in **Figure 24** for some indication as to the costs of not using CIDR address aggregation. Implementing CIDR can result in dramatic savings. For instance using CIDR all the hosts within one domain can be advertised as one prefix [Fuller *et a!* ,93]. Likewise, all the domains attached to one provider can be advertised as one prefix, as opposed to one prefix per domain. (CIDR is used in the NSFNET, but not is S.A.'s UNINET.)

❑**Benefits of the CIDR addressing plan**

Appropriate sized blocks of address space can be assigned without the low utilization of a class B address or the attached router overhead of multiple class C addresses. These address blocks can be custom sized and will typically be between 200 and 4000 hosts for mid-sized organisations. (See section 6.3.1) Pending the widespread use of CIDR capable routing protocols there should be a marked reduction in the growth of the routing table

sizes.

ClGrowth rate projections [Fuller *et al* ,93]

Growth rates for the Internet will be high initially as each service provider is allocated a contiguous block of IP addresses which should last them for at least 2 years. Once CIDR
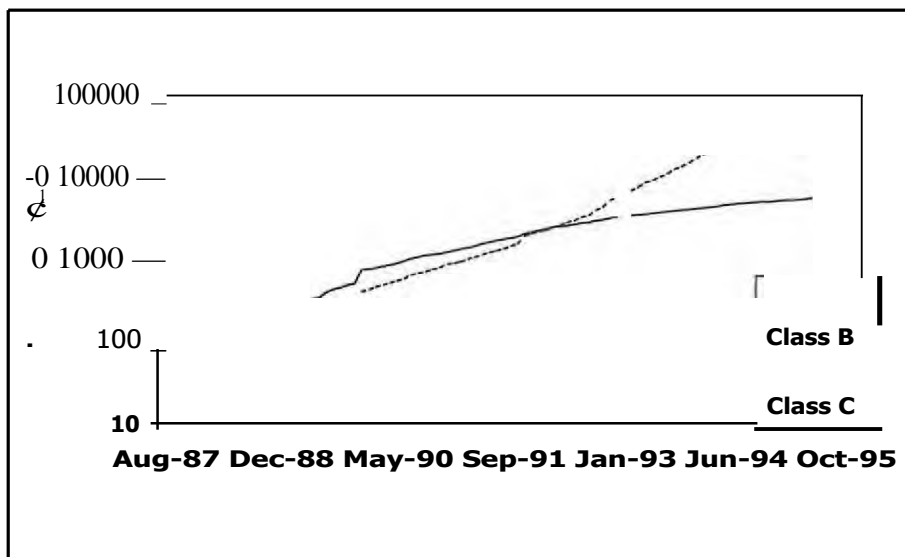
Figure 25 : Growth of Class B and Class C networks advertised by the NSFNET core routers.

is fully implemented the main growth factor influencing routing tables will be multi-homed organisations and their associated networks which have to be explicitly advertised. Consider the growth rate for Class B and Class C networks in **Figure 25.** This graph indicates clearly that the growth in Class B addresses was limited in the NSFNET core routers, possibly due to the introduction of CIDR at the inter-domain level. Likewise the growth in Class C addresses increased at the same time, further strengthening this assumption. (See section 2.2.1)

❑**Changes to inter-domain routing protocols**

CIDR requires changes to the way routing information is interpreted. Oid and new routing protocols will only interact successfully if supernetted information is exploded into individual network numbers (very memory intensive for the receiving system) or default routes (the suggested mechanism) are used. Domains running non-CIDR capable IGPs[n] cannot propagate supernet information through their domains.

❑**Semantic changes to routing protocols**

The most fundamental changes that have to be implemented are that routing destinations are now represented by a `<network prefix : CIDR mask>` tuple. Furthermore, routing is now performed on a longest-match' basis, and aggregation has to be supported by routing protocols in order to support CIDR.

The new rules for route advertisements are as follows:

1) Routing on longest match basis only, which means that multi-homed destinations have to be explicitly announced. This criterion guarantees consistency across routing algorithms.

2) Discard packets that match a summarisation (i.e. a mapping of multiple IP addresses into a single routing table entry) but do not match any of the explicit routes that make up the summarisation. This criterion is to prevent routing loops [Fuller *et al*, 93].

- Interior Gateway Protocol

[23] Longest-match routing - Routing is performed by finding the most specific route prefix which matches the destination address. E.g. the route prefix which matches the destination addresses best, i.e. has the largest number of matching bits, is always used.

ClAggregation responsibility

Responsibility for aggregation is in the domain that has had the block of addresses allocated to it. Normally this will involve the configuration of its border routers. The configuration mechanism can be performed either automatically or manually. Automatic configuration by dynamically learned routing information has the danger of not specifying precisely enough an address range when a route is not present. Automatic configuration is also unable to distinguish between a temporarily unreachable destination and an address that does not belong to an aggregate. Despite the advantage of not requiring manual intervention, the loss of flexibility in defining exact aggregate ranges makes manual preconfiguration a far more desirable choice.

❑ **Changes to intra-domain routing protocols**

Changes to intra-domain routing protocols arise because of often having to propagate external routing information internally for policy reasons or to aid routing, as with a transit routing domain. These changes can be performed in three different ways;

1) Using an Interior routing protocol which supports aggregation. This concept is easily implemented with OSPF or IS-IS which already treats routing information **as** destination+mask or prefix+prefix-length tuples.

2) Propagate exterior information without having to flood it internally.

3) Set up default routes for the discovery of paths to external destinations.

DiAddress space allocation considerations

CIDR requires that service providers are allocated an address space made up of multiple

class C addresses which fall on bit boundaries. To operate successfully InterNIC needs to change the address to name mapping scheme so that it no longer supports only octet boundaries in addresses. Providers should then allocate power of two blocks of IP addresses from the their own space to their subscribers.

7.1.4 Summary of the investigation into CIDR

CIDR solves the route scaling problem (as defined in section 3.2) very well, allowing for extra levels of hierarchy, without breaking the existing addressing scheme. Furthermore it allows for far better utilisation of the existing IP address space, giving some much needed time before this space is depleted.

It requires no topology changes, although implementing topological addressing and renumbering results in large savings. The main cost in implementing CIDR is in the requirements of new inter and intra-domain routing protocols. These costs can be offset by noticing that the process of upgrading the protocols is a gradual one and is effectively transparent to the user. Considering the advantages which bit-wise address masking has to offer it could be seen as a clear requirement for future routing protocols.

7.2 Address Extension by IP option usage (AEIOU)

AEIOU is a proposed life extension mechanism for IPv4 while awaiting IPng. AEIOU requires new host and router software which can make use of its proposed address extension mechanisms. Part of the motivation behind the AEIOU mechanism is that it should not break the existing IP routing protocols with its new addresses [Carpenter, 94].

7.2.1 What does AEIOU deal with?

AEIOU deals with the address space depletion problem, by implementing a temporary address extension scheme. AEIOU has taken a three-tiered approach by proposing the following steps before implementing its suggested mechanisms:

• No changes to the global Internet space or global routing architectures.

• Deploy CIDR[24] and BGP4 [25] whilst being very restrictive in new address allocations (cater for a maximum of 2 year growth). CIDR and BGP4 allow the use of better address aggregation and should hopefully slow the growth of the routing tables.

• No new allocation of network numbers to sites which already have network numbers of any class. These sites are to perform address reuse within their networks and apply CIDR where possible. The sites are also prime candidates for the implementation of AEIOU as will be seen shortly [Wang, 92][Carpenter, 94].

7.2.2 How does AEIOU work?

The exact address extension mechanism proposed by AEIOU works as follows:

---

[24] CIDR - Classless Inter Domain Routing - See Chapter 7 for more details.

[25] B GP4 - Inter-domain routing protocol which supports classless address masking, such as those used by CIDR.

- Define two new 1^3 option types, namely "Source address extension" (SAE) and "Destination address extension" (DAE)

SAE and DAE definitions

Type: [copied, class 0, option number TBD]
length: 6
data: 4 octet address extension (AE)

Figure 26 : Address extension definitions

These two options are defined as new IPv4 options with the parameters as shown in **Figure 26.** Each option carries with it a 32 bit (4 octet) extension which is assigned to internal hosts by each site when it runs out of its existing IPv4 address space. Using this mechanism, a host address within a domain will then become the tuple <IP address ; AEA , where the IP address will now be the site address and no longer a single host address. The site address is a single unique IPv4 address which the site chooses to be its AE prefix (Si = Site IP address). All inter-domain routing remains the same and so does the routing to all non AE hosts within the local domain [Carpenter, 94]. **Figure 27** illustrates an example of host addresses in two domains, the one using AEIOU, and the other not.
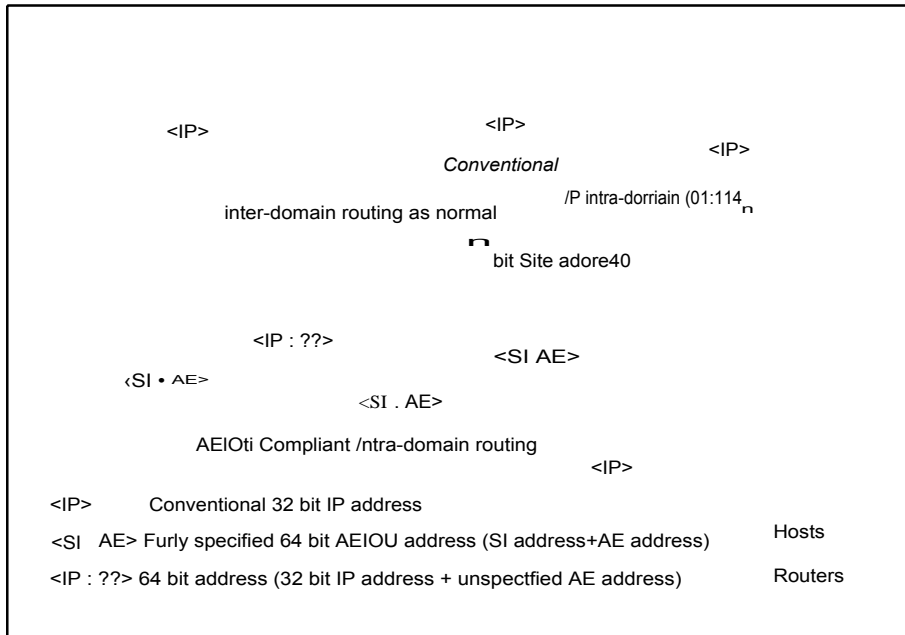
Figure 27 : Routing with AEIOU hosts in a local domain

Within the local domain there are now three types of hosts:

• Old hosts using 32 bit IP addresses only **<IP>.**

• New hosts using 64 bit <IP:??> tuple addresses, without an allocated AR These hosts are still within the globally unique 32 bit IP address space.

• AE hosts which have AE options implemented, and now have fully specified 64 bit addresses e.g. <SI : **AE>** tuple.

Communication between these hosts can only occur if both hosts can mould their addresses to have matching address structures. Eg. an old host can talk to a new host but not to a completely specified AE host. See **Figure 28** for the compatibility between the three AEIOU address strUctures.

	32 Bit LP address	64 Bit address (IP + unspecified AE address)	64 Bit address (SI + specified AE address)
32 Bit address (IP address)	*i*	*e*	*SS*
64 Bit address (IP + unspecified AE address)	*e*	*e*	*e*
64 Bit address (SI + specified AE address)	*S:8*	*I*	*e*

Figure 28 Which AEIOU hosts can communicate with each other.

The deployment plan for implementing the AEIOU mechanism is as follows:

1) All sites are allocated a unique site address (SI) from their existing IP address space. All routers are upgraded to support AE addressing. All clients are upgraded from old host addresses to new host addresses with an unspecified AE portion. At this point old clients will still be able to access new servers, which are using unspecified AE addresses.

2) All newly installed clients have to be AE hosts, which ensures that no more 32 bit IP address space is consumed.

3) Progressively convert old hosts using 32 bit IP addresses to AE hosts, by upgrading their software and renumbering them with 64 bit AE addresses .

4) Recover unused IP addresses which have been rendered unused by stage 3. These addresses can then be recovered and used in conjunction with CIDR for new sites.

5) Assuming there are no longer any old hosts, all servers can be converted to AE hosts and hence some more IP address space can be recovered [Carpenter, 94].

At this stage there will be a two tiered address space, the upper 32 bits managed via IANA or InterNIC whereas the lower 32 bits (AE) are managed at each site. See **Figure 29** for an illustration of a two tiered address space.
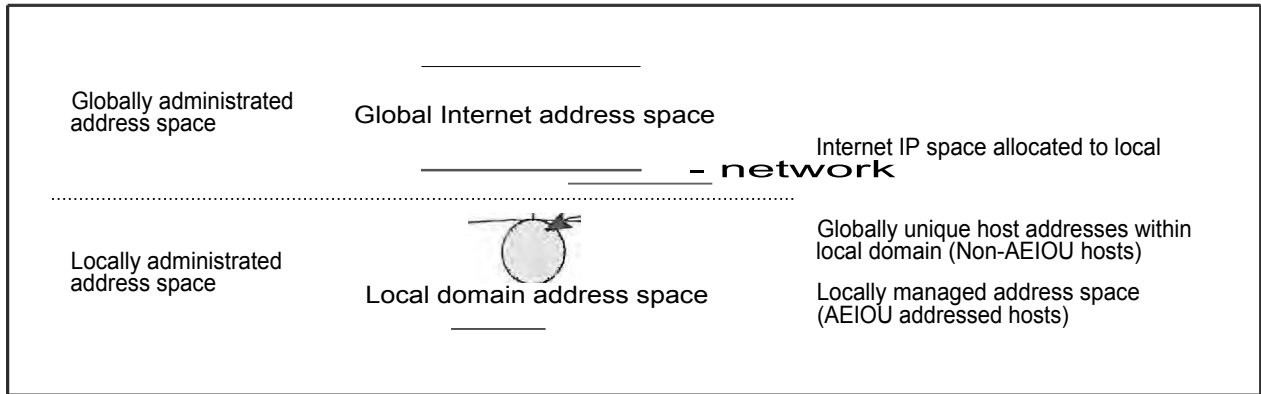
Figure 29 : Two tiered address space showing management responsibilities for address space under AEIOU.

The DNS server must return Routing Records which additionally indicate the appropriate address type for a host, namely old host (32 bit address), new host (64 bit address) or AE host (Completely specified 64 bit address). Likewise Intra-domain Routing protocols have to be modified to cater for AE hosts and have to look at the AE portion of the address tuple, instead of the IP address for further internal routing. There are a number of other small modifications which have to be effected to the general routing protocols to get AEIOU to work. Most of these modifications are to get the routing protocol to use the AE portion of the address from the IP option field in the packet header and not the IP source or destination address.

One might consider merely swapping the IP address and AE extension once the packet enters the intra-domain routing area, to circumvent the modifications mentioned above. This will not work as it violates the TCP and UDP checksum process.

7.23 Cost and Benefit Analysis of AEIOU

In analysing the AEIOU protocol, it becomes apparent that AEIOU requires extensive new software in both hosts and routers, which makes it very expensive to implement. New software implementation also has the side effect that it requires a very long period over which to be deployed.

Furthermore when considering the criteria in Chapter 6 Section 3, it becomes evident that AEIOU only tackles the address space problem and does not address the routing issue at all. The most serious problem in the Internet at the moment is that of the routing table explosion (see Chapter 3). Extended address space is unlikely to be of any use if routers can no longer route packets. Although AEIOU does indicate that it is to be implemented in conjunction with CIDR and BGP4 which will effectively take care of the routing scaling problem, this will nevertheless still be a very expensive temporary solution.

AEIOU deals with the sensitive issue of having to modify the various IP protocol layers such as the TCP layer by maintaining the existing 32 bit IP address. This in effect ignores the problem of the incorrect usage of the 1Pv4 address in the IP protocol (see Chapter 3) which it is felt, can only be detrimental to any future protocol development.

7.2.4 Summary of the investigation into AEIOU

As a short term solution to the IP addressing and routing problem, AEIOU is a very expensive solution. Any major software upgrading for the IP protocol is very expensive in terms of development and also takes a considerable amount of time to adequately penetrate the Internet. Being a short term solution, AEIOU does not address the issues relating to Type Of Service or policy routing requirements, nor in fact any other requirements for a new IPng. This can be interpreted as a problem since even an interim solution, were it to be adopted, is likely to have a life span of several years in the Internet. Thus not catering for a number of very necessary services or providing any appreciable motivation for user migration to AEIOU is a serious shortcoming.

7.3 NAT

NAT (Network Address Translator) is another protocol which uses the idea of a local address within the local domain and a global address in the global domain. Not each host has an address of both types which means that some sort of dynamic mapping has to take place.

7.3.1 What does NAT deal with?

NAT proposes a solution to the Internet addressing problem by allowing address reuse. The solution involves placing a Network Address Translator (NAT box) at the borders of the stub domains'. These NAT boxes have a number of globally unique IP numbers which they assign dynamically to hosts from within their domain requiring external connectivity [Tsuchiya *et al*, 93]. Default host addresses are only unique within their local domain.

---

[26] Stub Domain - Similar to a leaf domain in that it is a domain which only handles traffic for its own hosts.

7.3.2 How does NAT work?

❑ **A NAT example**

An example will now be presented where a host in stub domain A wants to send a packet to a host in stub domain B. Both stub domain A and B are implementing NAT, which
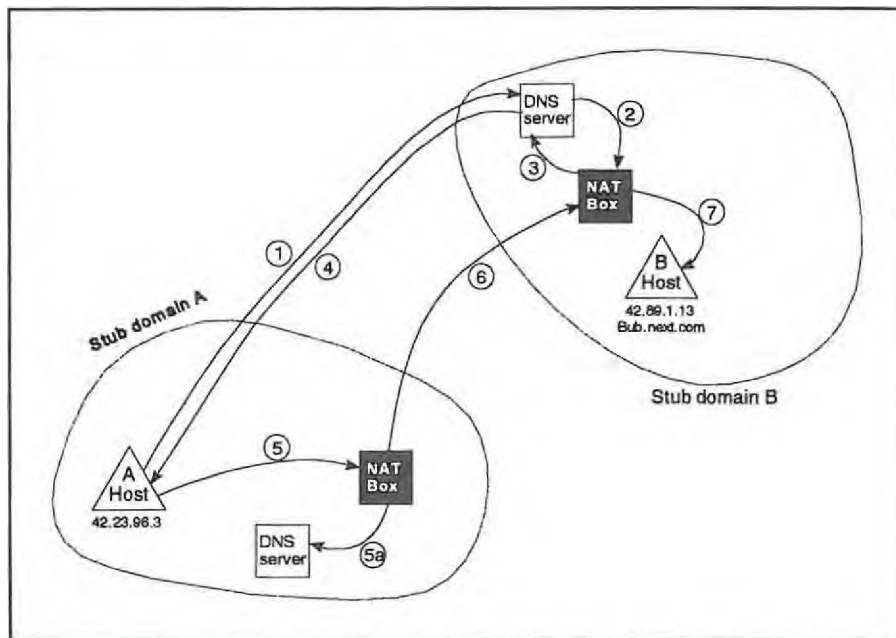
Figure 30 : NAT Host A in stub domain A sending a packet to NAT host B in another stub domain.

makes this the most complete case example for NAT.

The procedure for this operation is split into seven steps, as illustrated by **Figure 30.** Each of these steps which will now be explained in more detail.

© Host A (42.23.96.3) in stub domain A wishes to send a packet to host B (42.89.1.13) in stub domain B. Host A sends a DNS query to the DNS server in stub domain B, asking for the host B's (known to Host A as "Bub.next.com") global address.

® The DNS server in stub B knows Hosts B's internal address to be 42.89.1.13, but there

is no corresponding entry for a global address, so the DNS server sends an assignment request to the NAT box in its stub domain.

The NAT box in stub domain B finds an unassigned global address (146.231.128.57) from its address pool and assigns it to host B and returns this assignment to the DNS server.

C) The DNS server uses this newly assigned global IP address and responds to the original DNS query. The DNS server also updates the global address for host B in its tables, and will keep this address in its tables until it expires (i.e. expiration time elapses).

Host A then sends its packet to Host **B** with the destination address 146.231.128.57. When this packet reaches the NAT box in stub domain A, the NAT box assigns a global address (164.132.0.1) from its pool to Host A. At the same time the NAT box translates the source address in the outgoing packet to the newly assigned global address for host A (164.132.0.1).

© The packet is routed through the backbone to stub domain B, where it reaches the NAT box in this stub domain. The NAT box then translates the destination address (146.231.128.57) in the incoming packet to the local address for Host B (42.89.1.13).

IT The packet is then delivered to Host B.

❑ **Address Space and address assignment**

When using NAT the IP address space has to be partitioned into reusable addresses and globally unique addresses. The reusable addresses can be thought of as local addresses for use in the stub domains. The **IP** address partitioning should be complete and there should be no overlap, since this will lead to a case where a DNS can not distinguish between a local host or a host external to the stub domain.

An assignment scheme which allowed the use of a single class A address as the local address within all stub-domains and then allocated class B addresses as global addresses would be ideal. These class B addresses could be allocated to the regional backbones' and they in turn could subnet the Class **B** address and allocate a subnetted pool of addresses to the NAT boxes in each stub domain [Tsuchiya, Eng, 93]. Apart from providing good address utilisation, this scheme would also scale well as the regional backbone would only need to advertise a single network address for all the stub domains connected to it. This is illustrated in **Figure 31** where, a single Class **B** address would be advertised by the regional backbone.
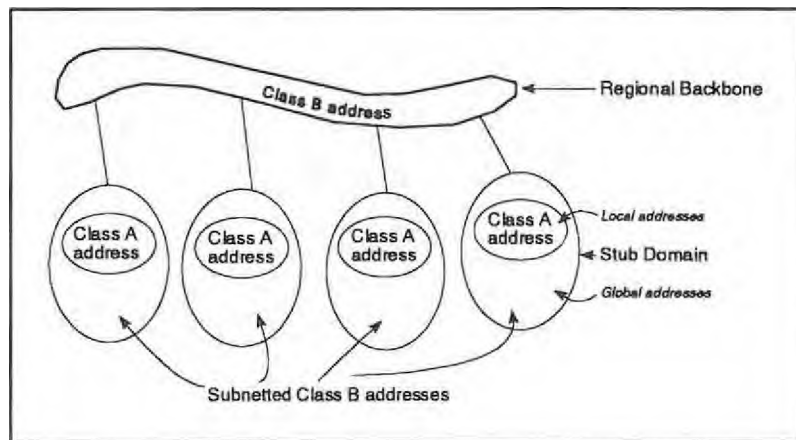
Figure 31 : Address allocation schemes when implementing NAT in stub domains.

Hosts which communicate outside the stub domain frequently, such as DNS servers, FTP servers and E-mail distributors could be assigned permanent external addresses. This provides a further subclassification of the global addresses used by NAT, namely:

• Static global addresses

     and

• Dynamic global addresses.

---

[27]Regional Backbones - Backbones which connect the stub domains to the rest of the Internet. (Also known as service providers)

The task of a NAT box maid ng an address assignment is a complex one, as it needs to make an assignment from within its global address pool which is not going to affect any existing connections. At the same time it must minimise the complexity and maximise the global address utilisation.

An *IP based* assignment algorithm which bases its activity purely on the timing of individual **IP** packets can be used, although it often results in destroying open connections which are merely idle [Tsuchiya, Eng, 93]. Alternatively some scheme of monitoring the connection status of traffic "flows"[28] could be used. This scheme could allow for partitioning of the global address space into two parts, namely global addresses which are involved in existing "flows" and global addresses where an end-of-connection indication has been sent (i.e. free to be reassigned again) [Hemrick, 85].

❑ **Routing with NAT**

The border routers of a stub domain have to intercept the routing data exchanged between domains and in conjunction with the current NAT address status make the appropriate changes to the local address information so that it reflects the correct addresses from the global address pool. Any other global routing information stays the same.

❑ **Different topology configurations with NAT**

When considering a stub domain, one has to realise that there may well be more than one NAT box and DNS server within a stub domain. This implies that a packet could potentially travel through more than one border router. If NAT is still to operate effectively, every NAT box and DNS server will have to know about the current dynamic global address assignments. In practice NAT boxes are grouped together into "cliques"[29]

---

[28] Flows" - this term is used to describe the traffic associated with a connection orientated protocol.

[29] Cliques - A NAT clique is a group of NAT boxes such that a packet addressed with an assignment from one of the NAT boxes in the clique can potentially be routed through any of the NAT boxes in the clique [Tsuchiya

which share information related to the current address assignments. Each NAT box still maintains its own pool of globally unique addresses but at the same time it also distributes its current address states to the other NAT boxes within its clique.
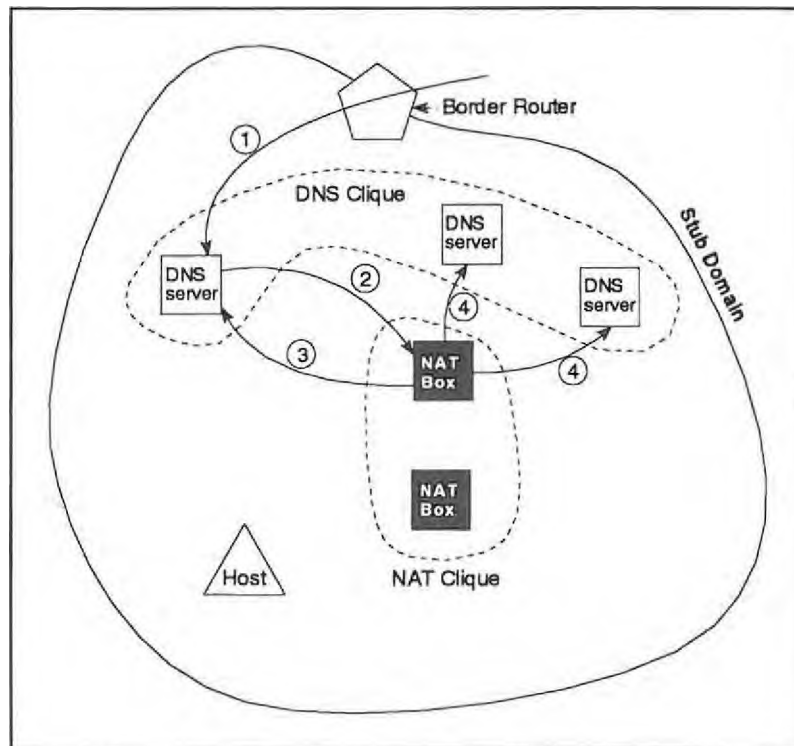
Figure 32 : NAT and DNS cliques.

The process which occurs when NAT cliques are used in a stub domain can be illustrated in **Figure 32** and works as follows:

CD DNS server receives a DNS query

2) DNS server sends a NAT-assignment query to a clique-assigning[30] NAT in each clique. (These queries are sent round-robin style to balance the load on the cliques)

---

[3]°Clique-assigning Nat box - a NAT box within a clique which is allowed to make global address assignments.

a Clique-assigning NAT receiving the query makes a global address assignment and returns the assignment to the DNS server. The assignment is returned with an expiration time[31] attached to it.

Clique-assigning NAT also sends an assignment notification message td all the DNS servers in the DNS clique

When NAT is used in a private network, which spans a backbone, performing NAT address assignment for each inter-organisational packet is very expensive. A technique used to combat this is encapsulation. Normally a number of static global addresses will be reserved for tunnelling between the different physical parts of the same private network [Tsuchiya, Eng, 93].

7.33 Cost and Benefit analysis of NAT

Some important points have been identified where NAT affects the Internet. By elaborating on these points, the author hopes to reinforce not only the importance of matching the specifications as discussed in Chapter 6, but also the importance of implementation and processing costs of a protocol.

Some of the specific implementation costs for NAT include modifications to the DNS servers and the border routers in the sub-domains. The operation of NAT is invisible to the existing IPv4 hosts within a stub domain. The modifications to the DNS server include adding code for querying the NAT box for a global address when the host required has only a local address. Border routers have to be modified to perform address substitution on incoming packets.

Some of the processing costs associated with running NAT in a sub-domain are: NAT requires some substantial header manipulations due to the constant replacement of **IP** addresses within the

---

[31]Expiration time - used to indicate when an address assignment can be removed from the address cache within the DNS server. (Similar in a sense to the TTL field within a packet.)

packet header. There are the direct address substitution into the packet header and the associated IP header and TCP checksum calculations which have to be recomputed after each substitution.

For ICMP messages there are a few further manipulations which have to be performed, not only on the header but also in the data portion of the ICMP packet. Since ICMP packets contain a portion of the header from the packet which resulted in the ICMP packet being generated, there has to be address substitution in this encapsulated header as well as the appropriate checksum modification as well.

Provided that the address allocation scheme as proposed in section 7.3.2 is adhered to, there is a considerable amount of address abstraction which can be achieved by using NAT in stub domains. Consider that if a regional provider has a single class B address and then subnets this address appropriately to all its stub domains, the provider need only advertise a single network address for all its stub domains.

In the case of FTP there has to be a modification to the FTP PORT Identifier, which includes the IP address. The port identification in FTP sessions is represented as an ASCII string which includes the IP address [Information Sciences Institute USC, 81b]. Upon substitution of a new **[P** address into this port identifier there is often a change in length of the port identifier (i.e in **Figure**

```
┌─────────────────────────────────────────────────────────────┐
│ Original Port identifier                                     │
│ <port number:IP address>          )000000c:1 0.18.27.3       │
│                                   ASCII length = Y           │
│                                                              │
│ New Port Identifier                                          │
│ <port number:IP address>          >oocococ:128.27,131.17     │
│                                   ASCII length = Y+3         │
└─────────────────────────────────────────────────────────────┘
```

Figure 33 : Address substitution for FTP sessions using NAT.

33 the port identifier changes in length from Y to Y-4-3) and hence the overall packet size, which in turn affects the TCP sequence numbers. See **Figure 33.** If the TCP sequence numbers are affected the **FTP** session will fail [Information Sciences Institute USC, 81a].

NAT can also result in an increase in misdelivered packets, since NAT cannot inform hosts when addresses have been re-assigned. This problem will result when DNS servers or hosts cache addresses longer than NAT boxes.

Lastly, any traffic which is using encryption at the IP level will automatically fail when trying to use NAT, as there can be no address substitution into these encrypted packets.

For every single dynamic global address change every DNS server in all the DNS cliques of a stub domain has to be notified and make changes to its tables. This house-keeping task is very time and processor intensive.

NAT has what may be considered to be an advantage, in that the address substitution prevents hosts from identifying each other. Despite this being considered an advantage in most cases it is a very large disadvantage for debugging as no tracing can be performed on packets.

7.3.4 Summary of the NAT protocol evaluation

After evaluating NAT against the criteria laid out in Chapter 6, the following conclusions can be drawn about the NAT protocol. NAT can be seen as a short term solution which attempts to solve firstly the addressing problem and to a limited extent the routing problem (route abstraction at the regional provider level).

The assumption made by NAT that most traffic in a network is intra-domain is one which will prove to be the Achilles tendon of this protocol. It was clearly seen in Chapter 2, that the Internet is experiencing unprecedented growth in areas which are very likely to involve inter-domain traffic to a large extent.

A further downfall of the proposal is the limited success when using FTP or any other application which makes explicit use of IP addresses.

7.4 Two tiered address structure

The two-tiered addressing approach deals with the addressing problem in the Internet by partitioning the address space in two. The one partition contains fully functional addresses, and the other contains addresses with restricted functionality.

7.4.1 What does the Two tiered address structure deal with?

The mechanism for this solution involves the use of two different address types, namely external and internal addresses. Both these address types have the existing 32 bit IP address format, with the exception that the internal address is only locally unique (within the domain).

The two tiered solution is motivated by the following assumptions:

• The majority of network traffic is confined within the domain, this being enforced by the nature of network applications and the limitations on inter-domain bandwidth.

• The number of machines operating as servers of some kind (i.e. DNS servers, ftp and mail servers) is far smaller than the number of machines in the entire domain.

• Most personal machines in a network are limited to local access within the domain.

• In many businesses only a few machines are allowed to be connected to the external networks (i.e. banks and government agencies) [Wang, 92a].

From the above assumptions the Two-Tier approach defines three classes of hosts within a domain, namely:

• **Servers:** those machines that are part of the computing infrastructure and require continuous external connectivity.

• **Isolated hosts:** those machines that are not allowed external connectivity.

• **Other hosts:** those machines that may require external connectivity, but not to such a degree as to require this connectivity all the time.

7.4.2 How does the Two tiered approach work?

The two tiered approach proposes that only the server class hosts within a domain be allocated permanent **IP** addresses. The remaining hosts that are allowed external connectivity are then to share a limited number of external **IP** addresses that are allocated to them by an External Address Sharing Service (EASS). The EASS is used to allocate a temporary external address to a machine which requests one. De-allocation of addresses and address updates within the DNS table are two of the other capabilities which are required when using the two tiered approach.

O Border Reuter   (1) DNS address request
A Host   (2) Address request reply
   DNS server   (3) Packet routed according to local address
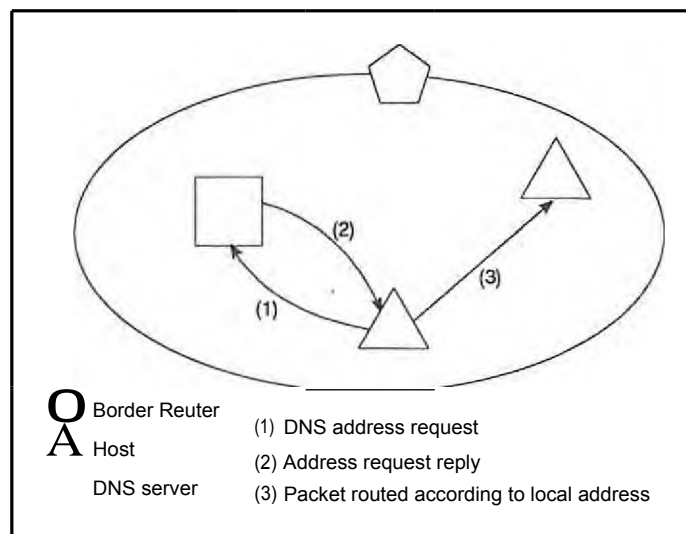
Figure 34 : Intra-domain routing using a local address.

Within the DNS tables there can now be two addresses for each host, one external address
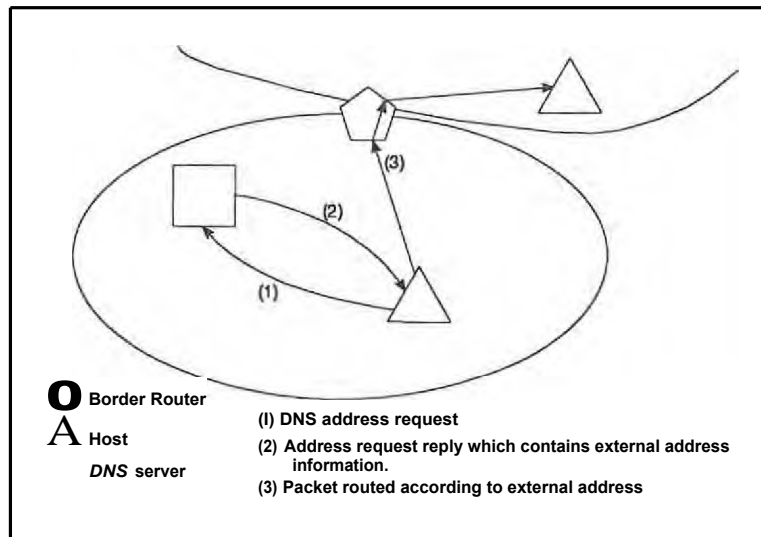
Figure 35   Inter-domain routing using an external address.

(sometimes present) and one internal address (always present). When servicing an address request, the DNS server must now examine the source and destination fields within the IP header and decide which IP address it is to return. In the case of intra-domain communication (see **Figure 34)** it returns the internal or local IP addresses, otherwise for inter-domain communication (see **Figure 35)** the external IP address is returned.

Allocation and de-allocation of addresses can be taken care of by a protocol such as DHCP[32] which will perform this process efficiently [Droms, 93].

Since there is going to be a change in the DNS table every time a temporary external address is allocated to a host it would be worthwhile considering co-locating the DNS server and EASS.

7.4.3 Cost and benefit analysis of the Two Tiered approach

This two tiered addressing mechanism was evaluated against the criteria in chapter 6. Some important issues relating to addressing and routing became evident as a result of this evaluation,

---

[32] DHCP - Dynamic Host Configuration Protocol, which manages host addresses dynamically.

and will be discussed in this section.
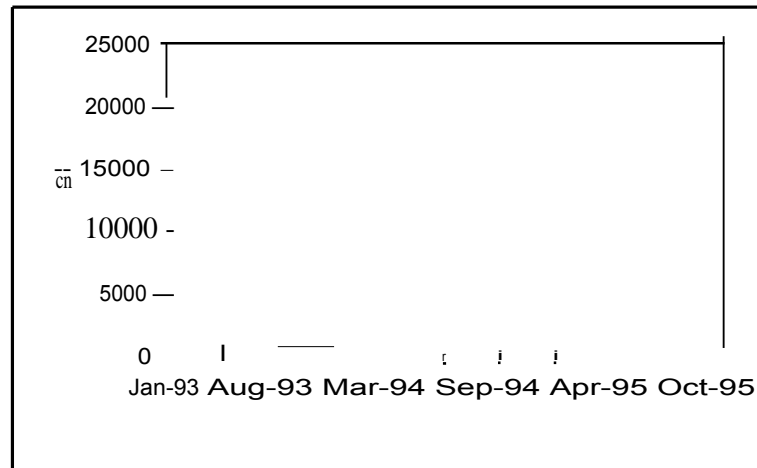
Figure 36 : Growth of the number of WWW servers in the Internet.

There is a major flaw in the assumption that most communication in a domain is intra-domain. The current trend in the Internet indicates exactly the opposite. The fastest growth area in the Internet is the World Wide Web (WWW) [Net Genesis, 95] (See **Figure 36)** which is an Internet-wide distributed system. With this trend the likelihood that most traffic is still going to remain intra-domain is very small. There may be a small number of commercial cases where the initial assumption holds true, but in general, public subscriber networks will not satisfy this assumption. Since the major growth areas within the Internet are in this public subscriber category, inter-domain communication is likely to increase greatly.

Existing IP address assignment:

Class A	126	nets	(7 network bits - 2 reserved network numbers)
Class B	16383	nets	(14 network bits - 1 reserved network number)
Class C	<u>2097151</u>	nets	(21 network bits - 1 reserved network number)

— 2113660 possible networks using only **IPv4**   [Pastel, 81a]

Using a two-tiered address structure each of these nets can accommodate
$$= 2^{32} \text{ hosts}$$

— Possible growth in the Internet using a two-tiered addressing system can support
- 2^{32} hosts * 2113660 networks
- 9×10^{15} hosts   [Wang *et al,* 92]

Figure 37   Potential growth using a two tiered addressing system.

The major benefit of a Two-Tier address structure is that each tier allows expansion to the extent where each network can use a large portion of the entire 32 bit address space within its local domain. The only disadvantage is chat only a small fraction of these addresses will be globally unique and thus be able to communicate externally without some sort of translation or mapping technique. Using the existing 32 bit address space and applying a two tiered address structure to it can cater for enormous growth as is illustrated in **Figure 37.**

The Two-Tier approach is an evolutionary approach which gives freedom to each domain as to when and if it wishes to implement the mechanism. It requires modification to **several** areas within the host software, but then again only for those hosts which are going to be using the EASS mechanisms. Since code modification is always expensive it should bring with it maximal benefits, and in the case of the two tier approach some benefits are present.

7.4.4 Summary of the investigation into the Two tiered addressing mechanism

The two tier approach provides an avenue for address space extension within the Internet. However, it does nothing to cater for routing table explosion problems, and is thus only a partial and temporary solution. It does offer some insight into mechanisms which could be used for address structuring and dynamic host configuration. Unfortunately the nature of the growth in the Internet will not tolerate a protocol which allows for host growth but at the same time explicitly prevents the growth of host inter-domain connectivity. This aspect alone excludes the Two Tiered approach as a serious contender even in the field of temporary solutions to the current Internet problems.

7.5 TUBA

TUBA (TCP and UDP with Bigger Addresses) is intended to be a long-term solution to the addressing and routing problems in the Internet. The proposal involves a gradual migration from TCP" and UDP[34] running over IP to TCP and UDP running over CLNP[35]. The long-term goal for TUBA is that NSAP[36] addresses should eventually replace **IP** addresses. TUBA intends to take a very gradual implementation path, during which time all existing IP operations should be unaffected [Callon, 92].

7.5.1 What does TUBA deal with?

By implementing NSAP addresses[37] TUBA intends to solve both the address depletion problem and the routing problem. NSAP addresses have a large number of formats and can have a maximum length of 20 bytes (160 bits). The basic outline of an NSAP address can be seen in
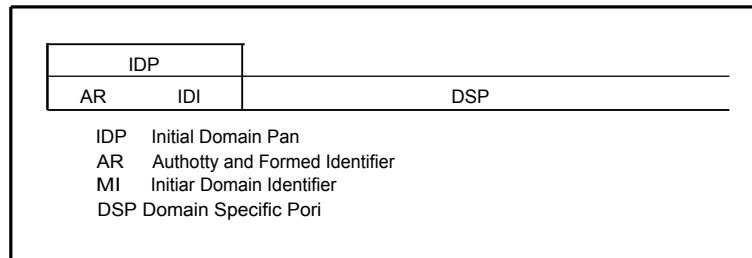
Figure 38 : Basic NSAP address structure

Figure 38 [Callon, 92][Piscitello, 93b].

---

[33]TCP - Transmission Control Protocol

[34]UDP - User Datagram Protocol

[35]CLNP ConnectionLess Network Protocol

[36]NSAP - Network Service Access Point

[37]NSAP addresses - "represent the endpoint of communication through the Network Layer and must be globally unique.¯ [Hernrick, 1985)

An NSAP address consists of two parts, the IDP or Initial Domain Part and the DSP or Domain Specific Part. The IDP consists of the Authority and Format Identifier (AFI) and the Initial Domain Identifier (IDI). The AFI specifies the format for the network addressing authority responsible for allocating values for the IDI. The semantics and the structure of the DSP are determined by the authority identified by the IDI.

The format which is proposed to be used by TUBA is most likely the NSAP address format for GOSIP version 2 (See **Figure 39).** GOSIP version 2 defines the DSP portion of the NSAP

IDP							
AFt	IDI	DSP					
47	0005	DFI'	AA	Rsvd	RD	Area	ID	Sel
1				2	2	2	6	1

Reid Sire 2n octets

IDP	Intial Domain Part	Rsvd	Reserved
AF1	Authority and Formal Identifier	RD	Routing Domain Identifier
1D1	Initial Domain Identifier	Area	Area Identifier
DSP	Domain Specific Part	ID	System Identifier
DFI	DSP Format Identifier	SEL	NSAP Selector
AA	Administrative Authority		
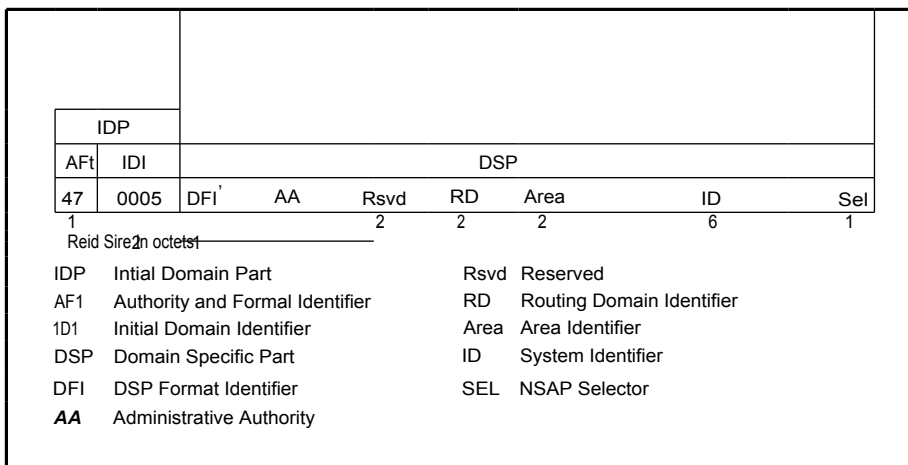
Figure 39 : NSAP address format for the GOS1P version 2 specification.

address as seen in **Figure 39.**

The two fields DFI (DSP Format Identifier) and AA (Administrative Authority) in the DSP define the authorities to whom the NSAP addresses apply. The fields which are of most interest to this work are the RD (Routing Domain Identifier), Area (Area Identifier), ID (System Identifier) and SEL (NSAP Selector) fields. These are the fields which are used for data abstraction and hierarchical routing. The RD and Area fields are defined to allow for better allocation of NSAPs along topological boundaries, which in turn will support data abstraction [Hemrick, 85].

A number of prefixes are defined for use with NSAP addresses and they are:

The Administrations Prefix is defined as: (AFT+MI+DF1+AA)

The Routing Prefix then becomes: Administration Prefix + (reserved') + RD

Finally the area address is specified as: Routing Prefix + Area.

Provided that the NSAP addresses are assigned from contiguous address space the above shown prefixes provide numerous levels within the hierarchy which can be used to implement address aggregation and data abstraction [Fuller *et al,* 92]. Furthermore, with the extensive hierarchy, the number of prefixes advertised at Backbone or TRD[39] levels will be significantly lower than a flat routing space [Collela, 91]. By implementing NSAP addresses with TUBA a multi-level hierarchical addressing space is ensured.

---

[38]Reserved field - The reserved Geld of 2 Bytes within a NSAP address is reserved for future use should added levels of hierarchy be required, or the addressing space become depleted.

[39]TRD - Transit Routing Domains otherwise known as service providers.

7.5.2 How does TUBA work?

Gl A TUBA Example

Consider the case illustrated in **Figure 40,** when an updated host (Host A) wishes to send a packet to another host (Host B) in another domain.
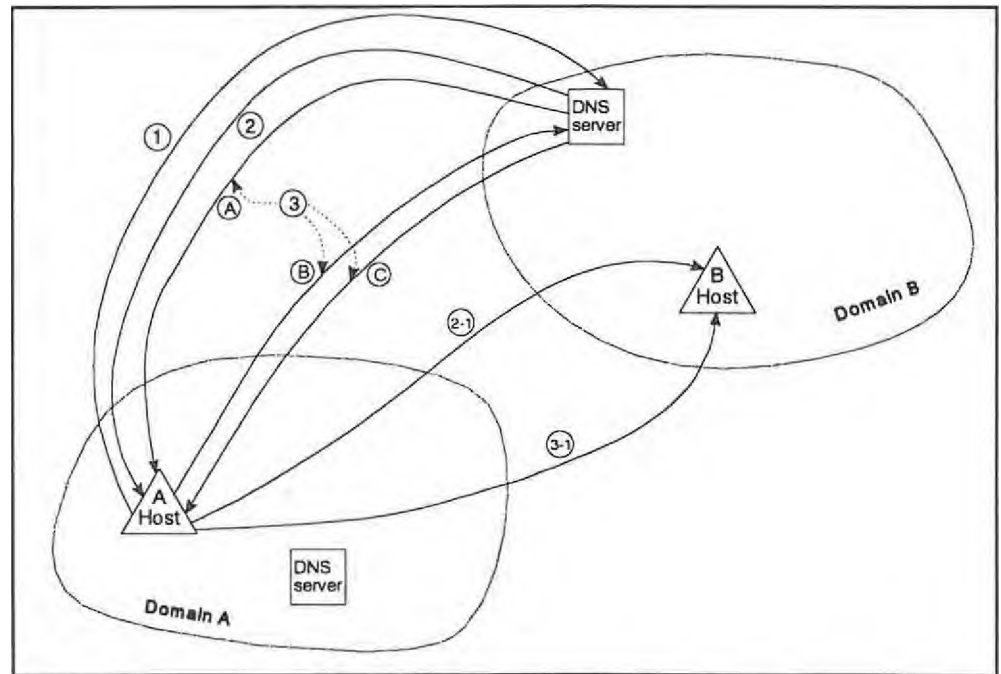
Figure 40 : Host A in domain A sending traffic to host **B** in domain **B.**

CD Host A sends a query to the DNS server in the domain of Host B, asking **for host B's long-address.**

e If both Host B and the DNS server have been updated to implement TUBA, and there are no routing limitations' between domain A or domain B, the DNS server returns both

---

[40]Routing limitations -Either the source or destination domain or intermediate domains cannot route CLNP packets.

the NSAP address and the IP address to Host A.

> (M- (D)The packet is then addressed using the NSAP addresses and routed
> according to CLNP [Piscitello, 930].

al Should either Host B or the DNS server in domain B be unable to support TUBA, the
DNS request fails (A) and the host in domain A responds by sending an old-style IP
address request to the DNS server in domain B (B), which responds with an IP address (C).

> (Z-11) In this case the packet will be routed using IP with a 32 bit IP address.

As can be seen from the example, TUBA requires extensive software modifications to
DNS servers, routers and hosts. DNS servers have to be capable of dealing with NSAP
addresses as well as IP addresses. Routers have to be capable of routing both IP packets
and CLNP packets. TUBA hosts have to know how to query the DNS servers for both
NSAP and IP addresses.

Modifications to the DNS servers

Updated DNS servers have a new type of resource record "long-address" which stores the
NSAP address for an updated host. Associated with the new long-address is a
corresponding name-to-address lookup "long-in-addr-arpa" (and reverse lookup) which
works similarly to the existing "in-addr-arpa" [Stevens, 94].
Inter-DNS communications can be performed using either CLNP or IP depending on the
availability information[41] and DNS server status (upgraded or IP only) for the surrounding
DNS servers [Piscitello, 93b].

Availability information - this is the information maintained within the DNS server, which indicates how it can
reach other DNS servers. For upgraded DNS servers this information will indicate an NSAP and IP address, whereas
for old servers it will indicate the EP address only.

❑ **Modifications to the updated hosts**

Updated hosts have to be modified to request a long-address first when trying to establish a hosts address. They also have to respond correctly to this initial request by either realising that the host or DNS server has not been updated, and hence sending a conventional IP address request, or receiving the NSAP and IP address from the initial request. Depending on the destination host's type (Updated host or IP only host) the source host must be able to generate either an IP packet or a CLNP packet. This means that TCP and UDP packets will be encapsulated directly into an IP packet or a CLNP packet depending on the destination host type. This encapsulation is run end-to-end which means that there is no CLNP to IP or IP to CLNP packet conversion at any point along the route [Piscitello, 93a].

❑ **Modifications to the updated routers**

Routers have to be capable of routing both CLNP and IP packets, which can be achieved by having the routers running dual protocol stacks. Until the routers in a domain have been upgraded there is no way in which TUBA can use NSAP addresses with CLNP. This is the primary upgrade dependancy for TUBA. Should hosts or DNS servers within the domain be upgraded prior to the routers, there will have to be some sort of configuration option which prevents the generation of CLNP packets and the use of NSAP addresses for entities in that domain.

7.5.3 Cost and Benefit analysis of TUBA

Some important issues which arise after evaluating TUBA against the criteria in alapter 6 will now be discussed. As with the analysis of the previous protocols, the emphasis will remain on addressing and routing issues.

TUBA minimises the risk associated with migration by allowing long term migration to an Internet where NSAP addresses and CLNP are the norm. The primary upgrade dependency is that routers have to be modified to be able to route CLNP. Upgrading the DNS servers and hosts will allow complete use of TUBA. Updated hosts in old style IP environments can be configured to generate IP packets and use IP addresses until their environments (DNS servers and routers) have been upgraded to support NSAP addresses and CLNP routing. (See section 6.1)

TUBA implements NSAP addresses which provide a large number of hierarchical levels within the addressing structure. To be able to fully utilise the inherent abstraction which can be gained from these levels, the NSAP addresses have to be allocated along topological boundaries. Effective address allocation will have to be ensured if the benefits of NSAP are to be gained. (See section 6.3.1)

NSAP addresses are very complicated [Collela, 91], and they may not be implemented well as a result of this. Furthermore, it should be borne in mind that at most only 15 of the 20 bytes actually contain information relating to routing or addressing[42], and even then a large proportion of the remaining high order fields are severely under-utilised. It remains to be seen if the complexity of the NSAP address is really necessary, and whether this complexity in fact benefits the Internet as a whole, or merely consumes resources.

FTP, NFS and other applications which use IP addresses internally or as part of their data will not be able to function properly with a larger address space, unless the applications are modified. This

---

[42] AR, IDI, Rsvd all contain information relating to internal NSAP address formatting or have large portions of their byte-space left unused.

in effect means that these applications will be limited for use to hosts and servers which still run old style IP addresses. Modifications to the software could be made to use the new larger addresses, although it would be a better choice to use an address independent but globally unique identifier such as the DNS name instead (See section 3.1) [Callon, 92].

TUBA can be used in conjunction with NAT once the 32 bit IP addresses fail [Callon, 92]. This would mean that the IP addresses would now only be locally significant within some stub-domain. A modification to NAT could be implemented whereby the packets exiting a stub-domain are mapped into CLNP packets. This means that all inter-domain traffic would eventually be via CLNP packets, with old style stub-domains running IP packets internally. This is illustrated in **Figure 41,** the explanation for the operation of these two protocols is as follows;
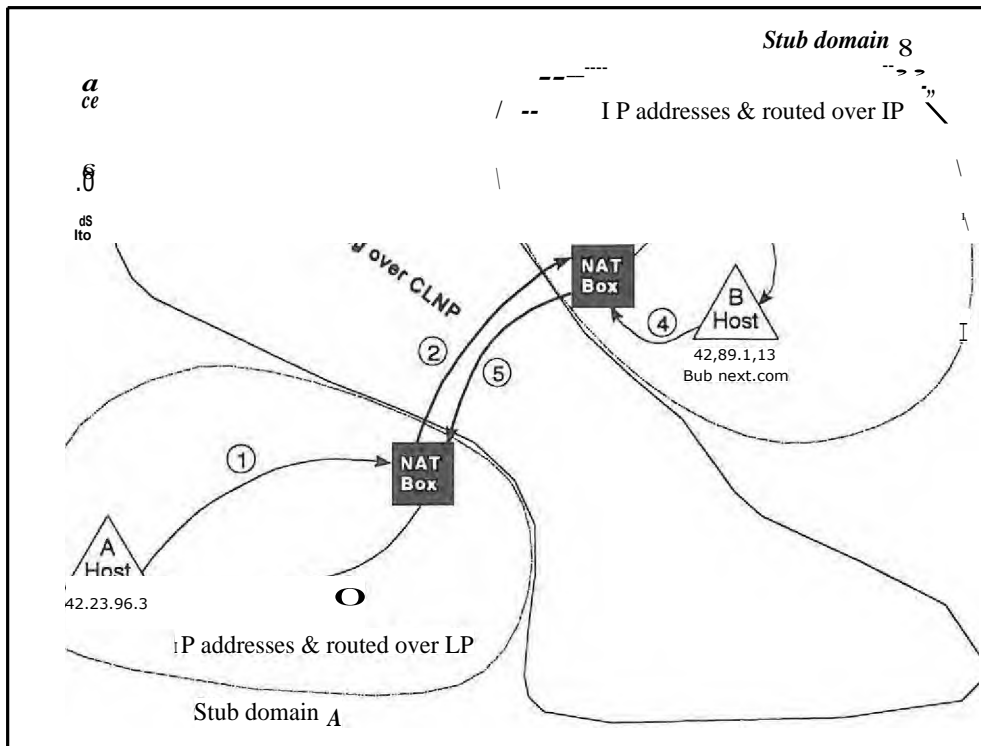
Figure 41 : TUBA with NAT implemented in the stub domains. Inter-domain routing done using NSAP addresses and CLNP with the intra-domain routing still using IP routing and addressing.

CD The datagrams are routed internally using the existing IPv4 addresses and IP datagram format.

At the domain border of stub domain A, the NAT box maps a temporary NSAP address to the existing IPv4 address and encapsulates the IPv4 datagram in a new datagram. This new datagram is routed through the TRD/backbone using the CLNP routing protocol.

The NAT box on the border of stub domain B performs another temporary address mapping between the NSAP address and a locally unique IPv4 address. The IPv4 datagram is extracted and routed internally using the conventional routing mechanisms to arrive at host B.

The process at 0, 6, and t is exactly the reverse of those illustrated in the first three points.

Using NAT has a number of costs associated with it, namely additional complexity and performance costs associated with the modifications required to packet headers and the encapsulation process. Furthermore, even when implementing NAT in conjunction with TUBA there still remains the problem of applications such as NFS and FTP which use IP addresses internally.

7.5.4 Summary of the investigation into TUBA

The idea behind TUBA is interesting and it does have some novel ideas such as running in conjunction with NAT. This scheme in a sense partitions the Internet into a dual protocol Internet, where all inter-domain traffic uses two protocols. The disadvantage of this is an excessive overhead for the constant packet encapsulation and unpacking.

TUBA requires the extensive re-writing of code in almost all of the entities within the Internet. In addition to this it does not solve the addressing problem in a neat or minimal way. Instead the use of NSAP addresses introduces a large and very complicated addressing mechanism, which does not bring sufficient benefits to warrant its use. Lastly, considering that there is a large amount of software which simply will not work under TUBA, it has to be discounted as a serious contender as a long-term solution to the problems of IPv4.

7.6 SIPP 16

SIPP 16 or "Simple Internet Protocol Plus 16" is a complete replacement protocol for IPv4. It is meant to be an evolutionary replacement which provides support for future functionality and at the same times maintains existing functionality.

SIPP 16 has a longer development history than most of the other protocols discussed in this document. SIPP 16 is the result of a merger of three separate protocols, namely IPAE, SIP and PIP [Crocker, 95]. Some of the best characteristics were selected from each of these protocols and incorporated into the original SIPP protocol. SIPP 16 was the result of upgrading the original SIPP protocol to use a longer address of 128 bits or 16 bytes.

The distinct characteristics with which each of the three original proposals influenced the final protocol can be summarised as follows:

• IPAE (IP Address Encapsulation) Dealt extensively with 32 bit IP address encapsulation into a larger address. Also looked at the associated transitional mechanisms required by this larger addressing scheme [Gilligian, 94].

• SIP (Simple Internet Protocol) was an upgrade on the existing IPv4 protocol which made use of 64 bit addresses, simplified datagram headers and provided additional extension headers to cater for header options [Deering, 92].

• PIP ("P" Internet Protocol) was a totally new protocol based on a new architecture. This meant that the architecture was not constrained in any way by the existing IP constraints. PIP contained a number of novel ideas such as variable length addressing, efficient forwarding using datagram flows and the logical separation of network layer addresses and host identifiers [Francis, 94c].

For the rest of this chapter SIPP and SIPP 16 are used interchangeably to mean SIPP 16.

7.6.1 What does SIPP 16 deal with?

❑ Address space

SIPP 16 defines a larger address space where addresses are 128 bits (or 16 bytes) long.
These addresses are strictly structured to provide hierarchical addressing within the
Internet (see Figure 43). The SIPP 16 address space is allocated to two main classes of

Allocation	Fraction of the address space
Provider-Based Unicast Address	1/4
Geographic Addresses	1/4
NSAP Addresses	1/64
IPX Addresses	1/64
Local Use addresses	1/256
Multicast Addresses	1/256

Remainder of the address space is Reserved for future use.

Figure 42 : SIPP 16 address space allocation (128 bit addresses).

A)

| Provider ID | Subscriber ID | Subnet ID | Node ID |

B)

| Unicast SIPP address |
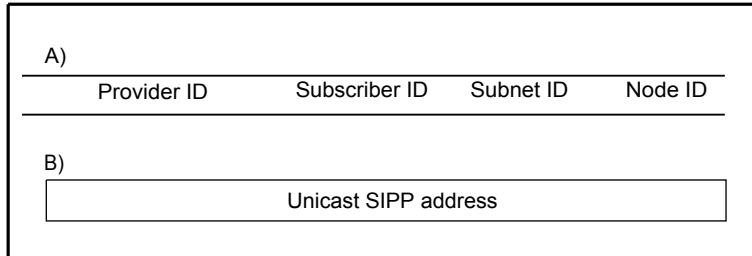
Figure 43 : Printer Server has limited knowledge of its address structure **B**
whereas the router has complete knowledge of its address structure A.

addresses, namely unicast addresses and multicast addresses [Deering, 94a] [Francis *et al,*
94d]. Unicast addresses are addresses that are used when a packet is to be sent to a
particular node. Multicast addresses, on the other hand, are used when a packet is to be
delivered to a group of nodes.

Other address types within the unicast address class which are catered for in S1PP 16 are:

Geographically allocated addresses, NSAP[43] and IPX[44] addresses and local use addresses. Local use addresses are for domains which have no Internet connectivity but which may one-day wish to be connected. By using local use addresses these domains prevent having to renumber completely when they become connected.

Hosts in the SIPP environment have differing levels of knowledge about the internal format of their addresses, depending on the function which they perform in their domain. For instance, routers will have more knowledge about their address format than terminals or print servers. The routers may know exactly what portion of their address is a provider prefix, subnet ID and node ID, whereas print servers may only be aware of a unicast address without knowledge of any internal structure. See **Figure 43.**

One of the SIPP 16 address formats which is of great importance for backward compatibility is the IPv4 address encapsulation scheme [Francis *et al,* 94d]. As can be seen in **Figure 44,** the IP address is simply encapsulated as the lower 32 bits within the SIPP address. The higher order bits then indicate whether the host can support SIPP in conjunction with IPv4 or not.

If SIPP is to solve the routing scaling problem the majority of its addresses will have to be

---

[43]NSAP addresses - Network Service Access Point addresses are used by the ISO protocols.

[44]IPX addresses - IPX addresses are used by the Novell (Novel Inc.) network protocols.

96 bits		32 bits	
0000	Prefix	.0000	IP address
		IP onry host	

96 bits		32 bits	
0000........,........	Prefix	0001	IP address
		I.	
		SIPP capable IP host	

Figure 44 : IPv4 address encapsulation inside a SIPP address. There is a bit flag to indicate if the IPv4 host is SIPP-capable.

allocated from within address space which can be aggregated efficiently [Gerich, 93] [Fuller *et al,* 93]. The two address space areas where this is likely to be the case are with Provider-Based unicast addresses and Geographically allocated addresses. A certain level of aggregation can be achieved using the SIPP address space as illustrated in **Figure 45.**

01	Provider ID	Subscriber ID	Subnet ID	Node ID
2	N tits	M tits	P bits	1213-(M+N+P)

Figure 45 : Levels of aggregation within a provider based SIPP address.

In this figure the node addresses within a subnet can be aggregated to a single subnet prefix, likewise all the subnets for a subscriber can be aggregated to a single subscriber prefix. Once again this aggregation can only be assumed possible if the address issuing authority assigns contiguous address space topologically. That is, all the subscribers for Provider **ID = X,** must be located within the same topological region and have the same provider prefix.

❏ **Routing Issues**

SIPP 16 uses bit wise masking (as proposed in CIDR) and the appropriate routing

protocols such as OSPF[45] which are CIDR capable. This scheme, which allows variable length subnetting coupled with the address hierarchy of a provider based SIPP address, has the ability to make SIPP datagram routing efficient. The provider based address space can result in an address space which aggregates well, and hence place low demands on the routing tables higher in the Internet [Deering *et al,* 94c].

❏**Datagram header optimisations**

SIPP 16 has redefined the datagram header, leaving out fields which were not used in **IPv4,** and making others optional to decrease the common-case processing cost of **SIPP** 16 datagram headers. The SIPP 16 protocol further defines the basic SIPP 16 header (see **Figure 46)** and the SIPP option header [Deering, 94b]. The basic SIPP header accompanies all datagrams through the network, whereas the option header is optional and is used to carry additional information relating to the datagram, such as hop-by-hop routing instructions or authentication information [Francis at al, 94d] [Hinden, 94b].

---

[45] OSPF - Open Shortest Path First is a link state dynamic routing protocol which supports subnetting, and hence ODR.[Stevens, 94]

Version	Flow label	
Payload length	Next header	Hop limit
Source address		
Destination address		

Version	Internet Protocol version 6
Flow label	28 bit field to Identify datagrams with special Type Of Service requirements.
Payload length	16 bit counter: number of octets of data
Next header	Identifies the type of header following the SIPP header.
Hop limit	8 bit integer which is decremented each hop (TTL)

Figure 46 : The basic SIPP 16 datagram header.

7.6.2 How does SIP? 16 work?

SIPP 16 is not radically different from IPv4. The datagrams are routed across the network using a link-state routing algorithm as opposed to a distance-vector' algorithm. This has the advantage of the network converging to a steady state quicker than when using a distance-vector algorithm [Stevens, 94]. The link-state routing protocol suggested, namely OSPF, has a number of other important advantages such as performing automatic load balancing and supporting multiple routes for each IP Type Of Service link [Stevens, 94]. There are other subtle changes to the protocols,

---

[46] Link- state routing algorithm - The data exchanged with neighbouring routers indicates the state of the current **links** between the router and its neighbours.

[47] Distance-vector routing algorithm - exchanges data matrices with its neighbours which contain some measure of distance between adjacent routers in the network, eg. RIP routing protocol.

but in general the DNS servers and hosts operate in a very similar fashion as with IPv4 [Information Science Institute USC, 81b] [Deering *et al, 94c*] [Thomson *et al, 94*].

The SIPP option header can be used to provide additional routing functionality within the network. Typical examples of what can be achieved through using SIPP routing headers are host mobility, provider selection and auto-readdressing (also known as host renumbering).

7.63 Cost and Benefit analysis of SIPP 16

The author analysed SIPP 16 against the criteria in chapter 6, once again paying particular attention to the items mentioned in section 6.3. SIPP 16 has a number of important addressing and routing implications, which will be explained briefly.

SIP? 16 solves the immediate problem of address space depletion (see section 3.1) by defining a very large SIPP address. This SIPP address space has been strictly divided into sections who's use has been predetermined. (E.g. SIPP encapsulated NSAP addresses are allocated 1/64 of the SIPP address space, with has a given prefix.)

Despite the SIPP 16 address space being so large and catering for a very large number of addresses, by defining its use so strictly there are a number of problems. Firstly the address space utilisation is likely to be very low, as large areas within the address space are going to remained unused. Secondly should there be a phenomenal growth in a particular type of address, and that particular address' space be depleted, it may be a non-trivial task to allocate some of the reserved space to that address type. This could be as a result of software being written to make explicit use of defined address areas only, hence excluding the reserved areas and making them very difficult to include at some later stage.

A SIPP 16 address performs two functions in the protocol. It identifies the interface' in the

---

[48] Interface - a node's attachment to a link (communication facility or the physical medium over which nodes communicate at the link layer)

Internet where a host is connected, and at the same time, uniquely identifies the host. This overloading on the term address is exactly the problem with the existing IPv4 addresses. (See section 3.3) Once a host decides to move it either takes its existing address with it and in so doing breaks the address aggregation, or it leaves its old address and has to be allocated a new address which it then advertises. The problem is that each host in the Internet needs some form of unique identification which is decoupled entirely from any particular routing function. (See section 6.3.1)

SIP? 16 provides added routing functionality by using routing headers. This provides the added functionality but in the process also increases the header overhead in datagrams.

The issue of implementation cost for a completely new protocol once again causes one the ask, whether the gains from the new protocol will adequately outweigh the disadvantages which the protocol implementation entails. In the case of SIP? 16 the answer is undoubtably yes, as it does provide some substantial advantages. The only doubtful area is the one of SIP? 16 having specified too exactly the utilisation of its address space. As pointed out earlier, there is very little way of knowing exactly what future requirements are going to be on Internet protocol suites (see section 6.2.1), and one would consider it a wise decision to allow for some flexibility in something as vital as addressing space.

7.6.4 Summary of the investigation into SIPP 16

SIPP 16 involves implementing a complete protocol, which is very expensive and has to undergo the associated debugging process. Since this is the case the rewards of implementing SIP? 16 must be worth the cost of implementation.

An area which the SIPP 16 protocol has not adequately addressed is that of catering for diverse requirements which future applications are likely to place on the Internet architecture. The SIP? 16 protocol has not provided enough flexibility in its routing algorithms or addressing schemes so that developers can augment the basic SIP? protocol with the necessary supporting mechanisms which may be required.

Clearly SIPP 16 solves the existing problems associated with IPv4 but whether it has more problems of its own remains to be seen, should it be implemented.

7.7 CATNIP

CATNIP or a Common Architecture for Next-generation Internet Protocol, is a proposal which provides a method for compressing the existing network layer protocols into a form which is identical irrespective of the original protocol. CATNIP has been designed to integrate the following network layer protocol formats: CLNP, IPv4, IPX (Novell Inc.) and SIPP 16 [Ullman, 94] [Mc Govern *et al ,94].* Using the CATNIP approach it will become possible for one end-system operating over a given network layer to inter-operate with another end-system operating over a different network layer.

7.7.1 What does CATNIP deal with?

CATNIP does not deal directly with the problems of IPv4, but instead provides an avenue for solving the problems indirectly. The common architecture which allows for integration between multiple network layer and transport layer protocols (see Figure 47), provides an avenue where a new IP protocol can be developed and implemented without affecting the existing IPv4 protocol at all_ CATNIP requires the use of a single consistent addressing method and common terms of reference to the entities within the system [McGovern *et al,* 94]. Being a long-term proposal
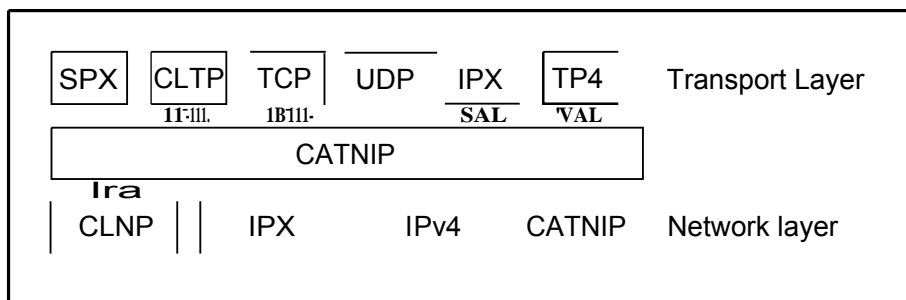
Figure 47 : CATNIP common architecture layer position in the protocol stack.

CATNIP can be used to facilitate the interaction between IPv4 hosts and any of the other supported protocols.

7.7.2 How does CATNIP work?

DICATNIP requirements

The two main requirements of CATNIP are a common network layer address format and a common network layer datagram format. The need for these two items is due to CATNIP's location within the protocol stack, namely between the network and transport layers. See **Figure 48.**

Transport Layer
CATNIP
Network layer
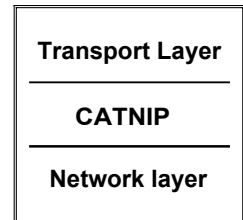
Figure 48 : CATNIP protocol layer position.

OA common network layer address format

The format of a CATNIP network layer address is given in **Figure 49. It** is similar to that of the OSI NSAP[49] address format [McGovern *et al,* 94] [Collela, 91]. The

Length	**AFI**	**IDI**	**DSP**
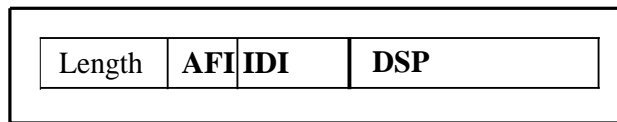
Figure 49 : CATNIP common network layer address format.

fields within the CATNIP address are named using OSI terminology such as AFI, IDI and DSP.

Each of these fields will now be explained in more detail:

Length   Number of bytes in the remainder of the address

AFI   Authority and Format Identifier, which determines the semantics of the IDI field. The AFI is a byte value which is registered with the ISO registering authority.

IDI   Initial Domain Identifier, which determines the authority for the remainder of the address. The IDI is a number assigned by the

---

[49] NSAP - Network Service Access Point, denotes a point in the network.

authority identified in the AFI.

DSP   Domain Specific Part, which contains the address as defined by the authority identified in the IDI field.

The idea with CATNIP addresses is that they can encapsulate the other network layer addresses using simple algorithmic manipulation [Ullman, 94]. For example, IPv4 addresses (32 bit) are encapsulated in the CATNIP address DSP field, with the appropriate values in the AFI *and* IDI fields to indicate that the encapsulated

Length	AR	IDI	DSP
7	192	AD num	IPv4 address

Figure 50 : IPv4 address encapsulation in a CATNIP address

address is an IPv4 address. See **Figure 50.**

IPX network layer encapsulation within a CATNIP address is illustrated in **Figure 51.** In the case of IPX the Internet address is the 32 bit IAN network number concatenated with the 48 bit MAC layer address. The resulting 80 bit address

Length	AFI	101	DSP
13	4_{4e}	Novell ICI)	Network+MAC address

Figure 51 : IPX address encapsulation in a CATNIP address.

becomes the content of the DSP layer in the CATNIP address.

SIPP 16 network layer addresses of 128 bits expand the CATNIP address to its full 20 byte length. In this case the DSP field stores the SIPP 16 address exactly,

Length	AFI	IDI	DSP
19	192	AD (0.1)	SIPP 16 address

Figure 52 : SUP 16 address encapsulation in a CATNIP address.

see **Figure 52** [Francis, *et al,* 94d].

CATNIP to OSI CLNP network layer address translations are simple as the address formats are exactly the same, and no encapsulation is required [Collela, 91].

O A common network layer datagram format

The format for the common network layer datagram can be seen in **Figure 53.** This network layer datagram is the result of the compression which CATNIP performs on the network layer protocols, to make each data unit look the same [McGovern *et al,* 94]. The datagram format is designed to cater for high performance networks but at the same time have a minimum header size, allowing it to be used in low bandwidth networks as well.

NLPID (70)	Header Size	D	S	R	M	E	MBZ	Time to live
Forward Cache identifier (FCI)								
Datagram Length								
Transport Protocol				Checksum				
Destination Address								
Source Address								
Options								
Data								

NLPID Network layer protocol identifier in OSI (Value 70 hex for IPv7)
Header Size   An 8 bit counter of 32 bit words in the header
Flags : 0 Destination Address Omitted)
       S Source Address Omitted)
       R Report Fragmentation done) Report if fragmentation was
           required for the datagram using an ICMP Datagram Too
           Big Message.
       M (Mandatory Router Option) insist that the router looks at
           the header options.
       E (Error Report Suppression)

Time to Live   B Bit count in seconds (TTL)
FCI   32 Bit cache identifier. (See text)
Datagram Length 32 Bit length of the datagram in bytes
Transport Protocol The original transport layer protocol
Options   Variable length options which may follow the header

Figure 53 : Format of a common network layer datagram, used by CATNIP

A streamlining mechanism for the processing of these datagrams has been provided in the form of the FCI (Forward Cache Identifier). The FCI is a 32 bit identifier which can be used to speed up routing decisions between successive routers. The FCI can also be used to maintain datagrams within flows, mobile host tunnels or circuits [Ullman, 94].

7.73 A CATNIP example

In the following CATNIP example a very simple case is chosen where the two communicating hosts are both IPv4 hosts. A second example will illustrate the communication between hosts using different protocols.

CI Example **1**
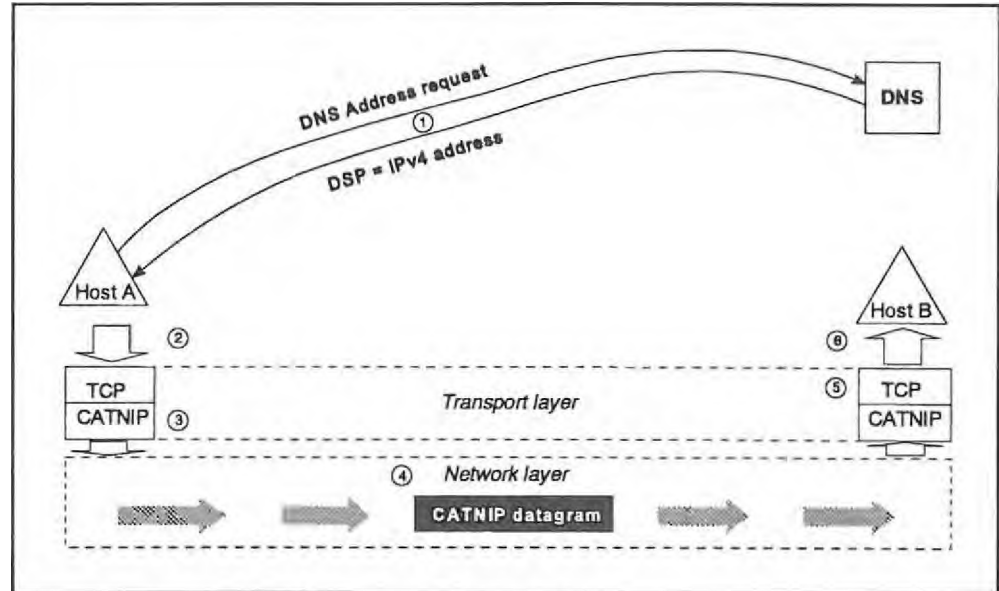
Figure 54 : Example 1, both hosts using the same protocols.

Figure 54 shows the interaction between Host A and Host B. Both Host A and B are IPv4 hosts. The process of routing a datagram from Host A to B can be explained as follows:

(it Host A queries Host B's address from the DNS server in Host B's domain. The address returned is in the NSAP format as defined by CATNIP, where the DSP field contains the 32 bit IPv4 address.

CD Host A generates an IP packet which is processed and moves down to the transport layer in the IPv4 stack,

® The datagram at the transport layer is translated into a CATNIP datagram, with all the appropriate CATNIP datagram fields set accordingly.

(4) Assuming CATNIP compliant routing, the datagram is routed to Host B's domain, and arrives at host B. The routing is performed using the full addressing form also known as

the NSAP address.

CID The CATNIP datagr,'rn is processed through the network layer and translated at the transport layer to emerge as an IPv4 datagram.

From this point the datagram continues up the IPv4 protocol stack as usual.

U Example 2
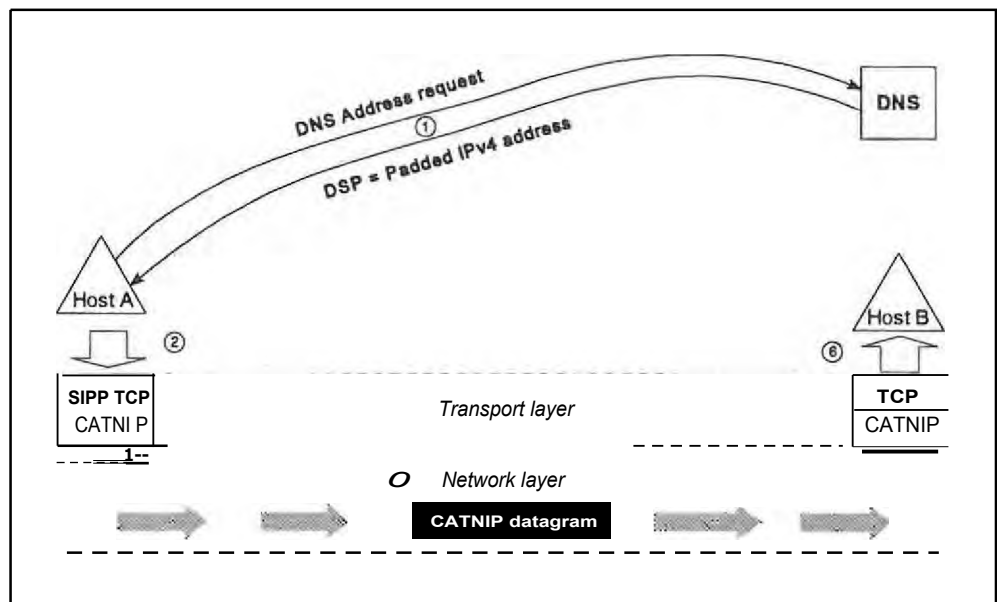
Figure 55 : Example 2, hosts using different protocols.

Figure 55 shows the interaction between two hosts which are using different protocol stacks (network and transport layer stacks). Host A is using the SIPP protocol stack, whereas Host B is using the IPv4 stack. For complete protocol interaction we assume both protocols to be using the full addressing form (or NSAP address).

O Host A queries Host B's address from the DNS server in Host B's domain. The address returned is in the NSAP format as defined by CATNIP, where the DSP field contains the

32 bit IPv4 address, which has been padded to match the full form addressing format (without this padding there can be no communication due to the incompatible addresses).

e Host A generates a SIPP 16 packet which is processed and moves down to the transport layer in the SIPP 16 stack.

Z The datagram at the transport layer is translated into a CATNIP datagram, with all the appropriate CATNIP datagram fields set accordingly. The translation involves most of the header fields, some of which are copied directly into CATNIP header fields (e.g. datagram header flags).

® Assuming CATNIP compliant routing, the datagram is routed to Host B's domain, and arrives at host B. The routing is performed using the full addressing form also known as the NSAP address.

Z The CATNIP datagram is processed through the network layer and translated at the transport layer to emerge as an IPv4 datagram. At this translation the datagram is translated from a CATNIP datagram to an IPv4 datagram.

8 From this point the datagram continues up the IPv4 protocol stack as usual.

7.7.4 Cost and Benefit analysis of CATNIP

After evaluating CATNIP against the criteria set out in chapter 6, some important influences of the protocol became apparent. These aspects as well as the implications of CATNIP on Internet addressing and routing will now be discussed.

CATNIP is completely incrementally deployable and has no deployment dependencies.

CATNIP provides stateless translation of network layer datagrams to and from CATNIP and hence by implication between the other network layer protocols as well. This provides a very desirable situation where there can be complete interaction between any of the supported network layer and transport layer protocols. Full inter-protocol interaction for IP and IPX can only be ensured once these protocols adopt the full addressing form'; up to that point these hosts will be able to communicate with hosts using their own protocols (See section 6.1) [Ullman, 94] [McGovern *et al*, 94].

CATNIP is efficient in its address translation process as it uses a purely algorithmic mechanism and there is no mapping or address hashing technique required. However, for two communicating hosts both using the same protocol (Protocol X), there is going to be a significant performance drop, since CATNIP is always going to have to translate the Protocol X datagrams into CATNIP datagrams, route these datagrams to the correct place, then translate back from CATNIP to Protocol X again. This performance drop is assuming normal hop-by-hop routing for the datagrams. Should a more efficient routing method be employed such as the use of FCI and datagram flows there may in fact be a performance gain, irrespective of the datagram translation.

CATNIP is not a solution to any of the IPv4 problems, (See chapter 3) but this does not make it unapplicable to the problem. CATNIP can be implemented in parallel to whatever network layer protocol is implemented for IPng and in so doing act as a facilitator for migration to the new

---

[50] Full addressing form - the NSAP format as defined and used by CATNIP as a common network layer address.

protocol.

7.7.5 Summary of the investigation into CATNIP

CATNIP provides a neat mechanism for protocol interaction which is very useful. As it does not cater directly for any of the IPv4 problems it can not be classified as a solution to any of these problems. It does however provide a very powerful mechanism for backward communication between IPv4 and IPng, which requires minimal code modification in either of these two protocols. By removing the burden of backward compatibility and inter-protocol interaction from the actual IPng protocol, CATNIP contributes to making the IPng design process easier. This means that IPng can be concerned with the direct issues and not have to worry about peripheral issues which can so easily detract from the overall protocol efficiency.

7.8 Nimrod

Nimrod is a complete new protocol which deals with the routing and addressing of datagrams for a very large internetwork.

7.8.1 What does Nimrod deal with?

Nimrod (A new IP Routing and Addressing Architecture) deals explicitly with the problem of routing which is currently facing the Internet. Nimrod identifies the current routing problem to be symptomatic of bad design with IPv4. The Nimrod proposal addresses the problem from basic principles, as follows;

How do we identify nodes in a very large internetwork and how do we get traffic from one node to another ?

An interesting view is taken by Nimrod designers in that it is clearly stated that addressing needs are secondary and must be adapted to suit the routing requirements [Chiappa, 93a]. This statement outlines the approach taken very clearly, namely that it is concerned with routing as the primary issue. Furthermore, the approach taken in the Nimrod design was one of designing a new protocol without the current constraints of what is being used and currently implemented. This approach should provide one with an optimal solution from where one can decide on a feasible new solution for implementation. The issue of migration was looked at once the final protocol had been selected.

7.8.2 How does Nimrod work?

❑ The Nimrod design approach

In the design process, Nimrod has isolated the following goals for the future Internet:

Olt has to be considerably larger than the existing Internet.

The size of the Internet is predicted to exceed even that of the current phone system in the near future. The Internet poses a number of other related and interesting problems which are not found in phone systems. Despite its predicted size it has to be made up of a very large number of autonomous systems and at the same time be able to operate in a very dynamic environment. The changes to the environment can be either topological or any number of other configurational changes [Chiappa, 93a].

*Maintain all the capabilities provided by the existing protocol.

Nimrod has to maintain all the current capabilities of IPv4. At the same time it must allow for incremental deployment i.e. require no change to host software to work (may require changes to take advantage of the added functionality). Nimrod does not insist that all nodes change from the current protocol, as there are too many unmodifiable hosts already. Any router changes need to be interoperable between IPv4 and Nimrod.

To encourage the use and deployment of Nimrod its Specifications must be complete and easily available as well as being vendor independent.

*Provide new capabilities which are currently a hindrance to the further development of the Internet.

A list of some of the current limitations of IPv4 which Nimrod aims to address are:

◆ **SIZE:**

The current protocol is running out of network numbers as well as address space. Due to its current size it is unable to route properly across the Internet.

*POLICY ROUTING[51];

The current mechanisms for policy routing are ad hoc and extremely complex. They are also unrelated to the desired goals, meaning that the mechanisms required to produce the desired results are complex and often impossible, i.e. they simply do not address policy requirements in any way. Policy requirements have three areas of importance, namely access control, trust model' and information hiding. Whatever policy control mechanism is implemented it must have the ability to add new policies in the future.

*TOPOLOGIES :

The current Internet has to have some sort of hierarchical structure to function. This does not represent the physical topology and hence is not efficient for routing or addressing.

+Attempt to provide necessities which will be required in the future.

Some of the future requirements for the Internet are that it be more immune to problems, be they accidental or deliberate. To develop effective solutions which counteract these problems, extensive simulation and test beds should be used. Provision for better security against deliberate attack is also required. The current technical task of configuration has to be minimised and automated to make the use of the Internet easier and hence more wide spread. In particular, the configuration of routing information in routers will have to be examined.

---

[51] Policy routing - external administrative input to the data routing process [Chiappa, 93a1.

[52] Trust model - determining which network elements can be trusted with policy sensitive traffic.
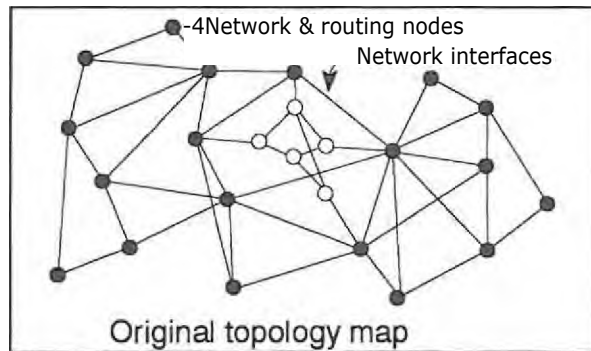
❑ **Nimrod architecture outline**

Figure 56 : Multi-level topological maps in Nimrod.

+Nimrod represents an internetwork as a **multi-level topological map.** Within this map are nodes and arcs, which express the available connectivity between the
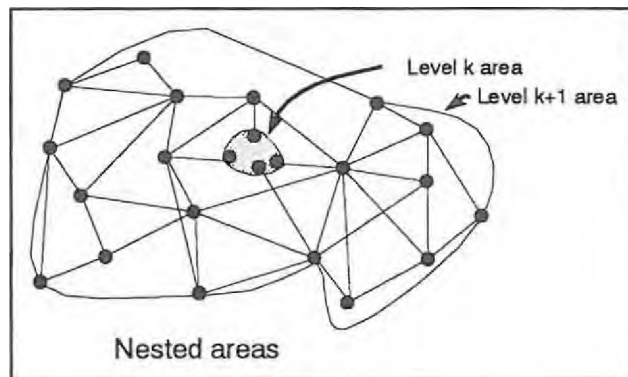

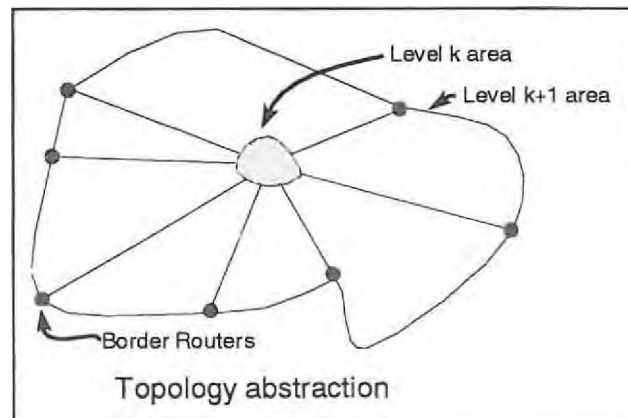**Figure 57 :** Abstraction using areas

Figure 58 : Nesting abstracted areas in the topology

different nodes in the intemetwork. Each node identifies a region within the network. This region can be as small or as large as is required. The arcs within the map represent unidirectional traffic flow between the nodes at the head and tail of the arc. See Figure 56. Each node within the internetwork has a number of attributes relating to its connectivity. These attributes include characteristics such as: which arcs[53] are connected to the node, internal map characteristics[54], transit connectivity" and lastly inbound[56] and outbound[57] connectivity.

The multi-level characteristic is present due to the method of abstraction which is used, namely the idea of areas. Areas are subgraphs within the map which can be abstracted using some criteria. See **Figure 57.** These areas can be nested and additionally each area can be given a level attribute for illustrative purposes. This level attribute is always one higher than the highest level attribute nested within that area. See **Figure 58.**

The reason for introducing the idea of areas is to enable topology abstraction, which means that an area can be replaced by an equivalent representation or some simpler representation for the actual topology. Since the method of abstraction is flexible within Nimrod, any number of abstraction mechanisms can be used. The abstraction for an area is determined by the area's border routers, which implement the abstraction mechanism and advertise the abstracted area to the rest of the internetwork. Abstracted areas can be represented in the topology map as one or

---

[53] Arc attribute - indicates which neighbouring nodes are connected to the current node.

[54] Internal map characteristics - a representation of the internal connections within a node. This characteristic can be nested and can also vary greatly in complexity.

[55] Transit connectivity - enumerates the services available between adjacent sources and adjacent destinations.

[56] Inbound connectivity - a map of inbound connections into the node, when a complete internal map is not available for route calculation.

[57] Outbound connectivity - presents all the connections between points within the node and adjacent destination outside the node.

more nodes with internal map attributes.

A number of abstraction mechanisms have been proposed, each of which will be described shortly.

*The first method involves replacing an area with only its border routers and a pseudo node which has the same characteristics as the real topology for that area. See **Figure 59.** The policy provided to the outside world would then depend on which pair of border routers were chosen for the datagram route.
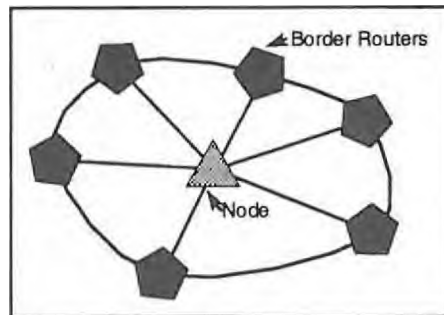

Figure 59 : Pseudo node abstraction technique

♦ The second method involves replacing an area with all its border routers
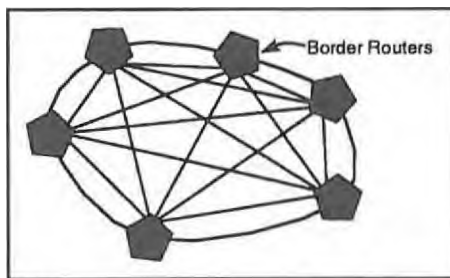

Figure 61 : Fully connected border routers
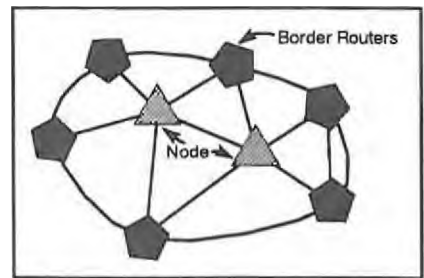
Figure 60 : Customised abstraction

and have these border routers fully connected as can be seen in **Figure 61.** This approach has the advantage of being able to present a more accurate

link-characterising metric to the outside world.

*Alternatively user interaction could be required for some form of user-customised abstraction as is shown in **Figure 60.**

Most of the abstraction mechanisms make use of the idea of replacing a number of physical links with one or more virtual links. These virtual links can be defined as the links made between two border routers. They can be advertised with link attributes which are calculated from the physical link's attributes.

0 The identification of users of the internetwork layer is achieved in Nimrod by defining an End Point Identifier **(EID).** The format of an EID is a variable length bit string, comprised of 64 bit components. An EID has no topological significance what so ever and is the identifer by which every host in the Internet is uniquely known. **To** provide a more user-friendly version of the EID, an Endpoint **Label (EL)** has been defined. The EL is an unlimited length ASCII string which is structured hierarchically and used as a key into DNS tables. This scenario is very similar to the current IP number and machine name identifiers used with current DNS tables. The relationship between EIDs, ELs and hosts within a system can be illustrated in **Figure.62.**

End Poinl Identifier	**DNS RECORDS**	**End Point Label**
F1BACD001A7BEI001 $_{16}$		Cs-adwin.hosl01
Machine hien* name		User friendly name

Figure 62 : EIDs and ELS in Nimrod

Nimrod makes user of another concept, namely that of a Basic Topological Entity (BTE). A BTE is used to distinguish between the components of a map, namely nodes and connectivity specifications.

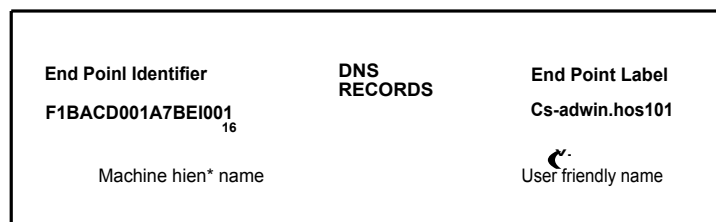Accompanying the multi-level topology in Nimrod is the multi-level address or multi-level locator. In Nimrod an address or locator is strictly a point within the topology map. It is non portable and indicates an attachment point or BTE in the network. Care has been taken in Nimrod to use an addressing scheme which best matches the representation of the network topology. This is significant as routing is performed using locators only.

The general format of the locator is a multi-level hierarchy with a variable number of levels of variable length. The exact format of a Nimrod locator can be seen in Figure 63. Locators have the property that they own any locators which have their locator as a prefix. E.g. Locator a:d owns the locator a:d:c since a:d:c has the prefix a:d.

Locators satisfy the requirement of being topologically sensitive by insisting that locators which share the same prefix be located in a contiguous region of the network topology map.

Add_Element : Add_Element $_i$:   : Add_Element $_3$:Add_Element

Where each Add_Element,ix=i I) is a unique identifier for some node within the network
and Add_Element $_p$ is the unique identifier for some real physical asset.

Figure 63 Nimrod locator structure

An EID is not permanently coupled to the Nimrod locator in any way and mapping between the two is flexible and depends only on the user's current location.

Initially the existing 32 bit IP addresses can be used as EIDs by padding them with Os and still maintaining uniqueness. Since there is no internal structure to an EID, all the numbers within the 32 bit IP address space can eventually be used, approximately $4x10^9$ unique numbers.

❖Nimrod supports four different datagram forwarding mechanisms:

Connectivity Specification Chain forwarding (CSC), Connectivity Specification Sequence forwarding (CSS), Flow mode forwarding and datagram forwarding. Each of these will now be considered n more detail.

*Connectivity Specification Chain forwarding (CSC):

CSC forwarding can be considered to be a derivative of Strict Source Routing in IPv4. It operates by having a list of locators carried in the datagram header. These locators correspond to connectivity specifications
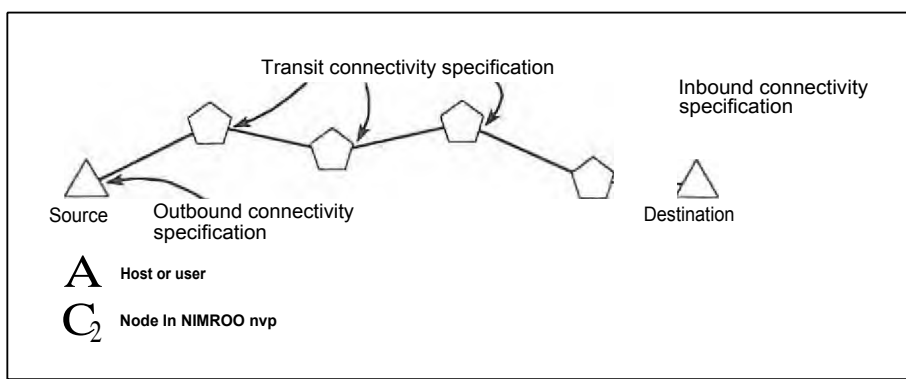
Figure 64 : CSC forwarding mode in Nimrod

which are node attributes. Routes specified to use CSC forwarding are initiated by specifying connectivity specifications and not physical entities. The locators in a CSC header would typically correspond to a type of service between two users in the intemetwork. There should be a direct link between any two consecutive connectivity specifications in the datagram header, as the datagram has to be forwarded strictly according to the locators (connectivity specifications) specified. For an example of **a** CSC path see **Figure 64.**

*Connectivity Specification Sequence forwarding (CSS):

CSS forwarding is analogous to Loose Source **Routing** in **IPv4. It** operates almost exactly as CSC forwarding with a relaxation on the contiguity

requirements for consecutive locators.

*Datagram Mode forwarding:

Datagram mode forwarding is an optimisation to cater for datagram traffic. It is the Nimrod equivalent of hop-by-hop routing in IPv4. Nimrod uses a pre-set flow state in all the routers for this forwarding mode. The routing
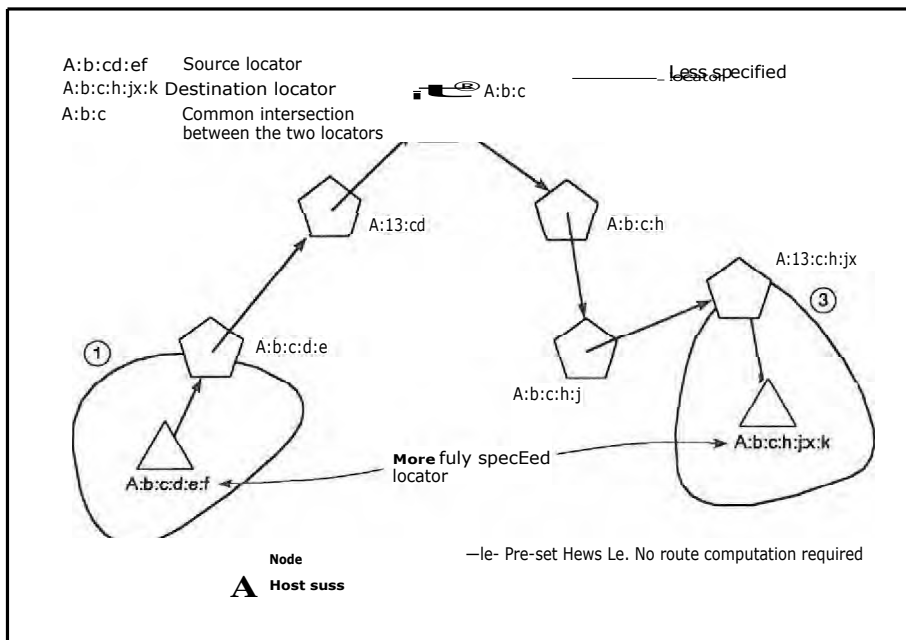
Figure 65 : Datagram mode forwarding in Nimrod, illustrating the routing mechanism using less specific and more specific locators.

mechanism for datagram forwarding is as follows (see **Figure 65** for the illustration):

CD The locator is completely specified at the host level (Le it identifies a single host). Each active router selects the appropriate flow to send the datagram to the next higher level object (less specified) within the source locator.

® This process is repeated until the current object has an intersection

(shares a common Nimrod locator) with the destination locator (At this point the locators are the least specified i.e. the given locator is a prefix to a great number of hosts and networks.).

After this stage the routers then route the datagram according to the next lower object in the destination address (The locator becomes more specified after each router to the point where it is specific enough to specify a single host in the network (i.e Fully specified locator)).

There may be cases where the pre-set flow for datagram routing does not exist. In this case the datagram is sent in conjunction with a flow setup request to the appropriate object. Flows can be established to make the routing of datagrams between two nodes within the internetwork more efficient. This can be achieved by establishing a flow which avoids having to traverse up the source destination locator and then down the destination locator. See **Figure 65** [Castineyra *et al, 94]*.

♦ Flow Mode forwarding:
 A flow can be considered to be a virtual path for datagram transmission across the internetwork. These flows may be set up in a variety of ways, either having their state stored in the intermediate routers or carried in every datagram which is part of the flow. Nimrod chosen the latter where a flow is set up by explicit definition in the routers and then recognised later by a globally unique flow-tag mechanism. This tag mechanism (Flow Identifier field) is **all** that is carried in the packets once the flow has been set up. For the purposes of this discussion we are concerned with what happens to a flow when there is flow partitioning within an area. If there is local component failure within an area (i.e. failure between point A and **B** in **Figure 66**) a flow may be able to be repaired by the area's border routers (i.e. re-route the flow via C between A and **B** in **Figure 66**)
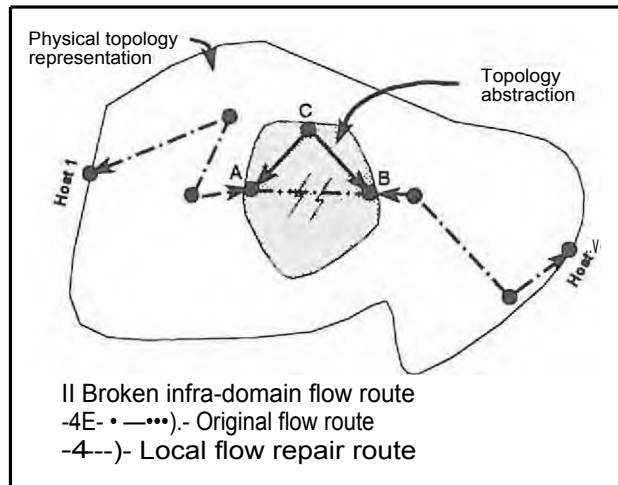
Figure 66 : Local area flow repair

without notifying the flow initiator or first hop router (source route generator). Should this not succeed, the flow repair task can be passed up to an area one level higher and repeated. This entire process can be executed recursively until the flow is repaired. The only condition under which the source router generator is notified of a flow breakage is when an asset explicitly selected by the generator has failed and the rerouting of the flow has to be determined by the source route generator and can not be taken care of invisibly by the network.
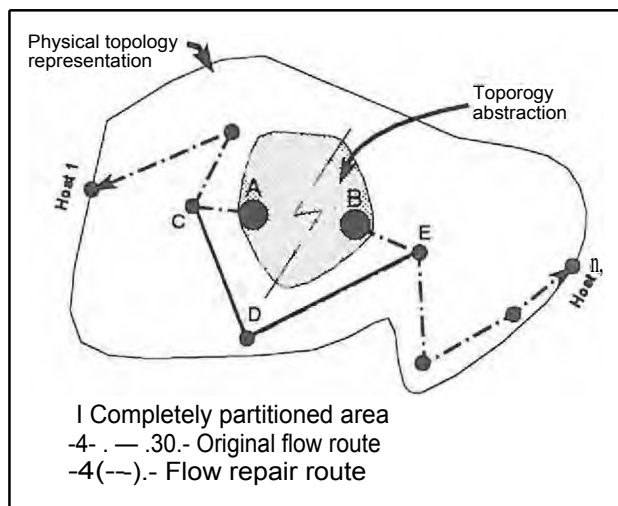
Figure 67 : Complete domain partitioning and flow repair

A similar mechanism is employed for repairing flows when there is complete area partitioning. In Figure 67 there is complete area partitioning between A and B. The flow is then rerouted around the partitioned area through C, D and E. The rest of the flow path remains unchanged.

*Nimrod has split the routing process as is currently used in the Internet into two separate processes, namely routing information distribution and route calculation. The routing **information flooding** is performed by the routers and the mechanism is specified within the Nimrod protocol. However, since routing mechanisms are likely to receive much attention and development in the future, Nimrod has provided an interface for routing mechanisms, and not explicitly specified how the routing is to be done. This allows the best routing mechanism for the task to be implemented in conjunction with Nimrod. At the same time not fixing the routing mechanism also means that as new, better and faster mechanisms are developed they can be deployed without added protocol changes.

The routing algorithm used by Nimrod is a multi-level Link State Algorithm. The algorithm is multi-level to be able to handle the large amounts of routing state which are in the multi-level system. State information within the system relates to connectivity information, routing policy information and Type Of Service requirements of each node. As mentioned earlier this state is stored as attributes for each node within the topology.

Using a Link State architecture does mean that the issue of topology abstraction has to be considered in more detail. Since the abstraction mechanism chosen is exterior to the Nimrod protocol it allows for input from administrators as to which method they feel to be best for their environment. Non-optimal routes or un-found

---

[ss]Routing state - information pertaining to the route and the policies required for that particular route.

routes are the result of using a pure LS algorithm. These two problems can be overcome through the use of Source routing in the form of CSS, CSC or flow mode 'orwarding. Abstraction is performed locally at each border router so that the level of abstraction can be decided on by each route generating agent.

Source routing is a necessary addition due to the policy routing considerations and the expandable attribute list for links, since policy considerations for a route are specified at the initiator. Using source routing allows for the incremental deployment of better algorithms for route finding and creation, as these mechanisms can be implemented on the first border router and then used when routing through it.

Using source routing does not mean that every packet has to carry the complete data for setting up the route. Once the route has been set up it is very unlikely that any on-the-fly modifications will be allowed. The benefit of carrying the set-up data in each packet is that the routers can remain relatively stateless, although it is more likely and feasible for a route set-up stage to exist whilst a flow is established. After this point flow modifications by routers will be subject to the constraints as set by the flow initiator. For this purpose routers which do not understand a routing attribute must be able to decide if the attribute is a restrictive or informational attribute. Since the entire route is calculated by one router it does not matter if different routers use different routing algorithms.

$$Linlc_{,ic} = \begin{array}{ll} metric_{mTu} & * \quad weight_{\ y} \\ metric_{Bandwidth} & * \quad weight_{Bandwithh} \ + \\ metric_{Latency} & \quad weight_{Latency} \\ & * \quad ....+ \\ metric_{s} & * \quad weight_{s} \end{array}$$

Figure 68 : Composite Metric Calculation

A sample routing algorithm which can be used by Nimrod is to calculate a routing tree by walking through the map and dropping all links which do not satisfy the routing policy or Type Of Service requirements. This map can then be minimised according to some constraint using a Dijkstra minimisation algorithm. The constraints could be a composite metric of different link attributes or policies, where the metric is computed by giving each attribute or policy a weight, as can be seen in **Figure 68.**

7.83 Nimrod example

We will consider an example which deals with abstracting a physical topology map into a Nimrod topology map. The different levels of abstraction will also be illustrated.
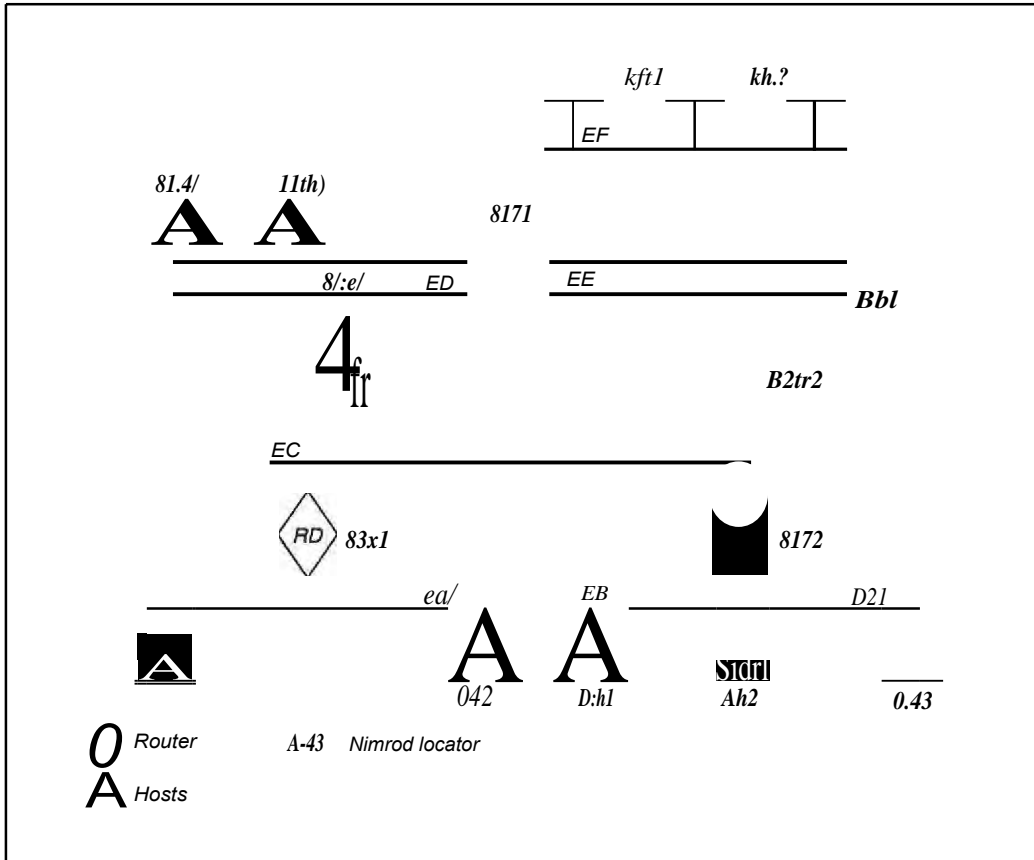
Figure 69 : Physical network topology for Nimrod example.

Consider the physical network as shown in **Figure 69 .** This network consists of six different networks **EA.. EF,** connected via routers **RA..RF,** containing hosts **HA..HK.** The labels in capital letters are the entity names with the Nimrod locators represented in their locator:locator format.

As was discussed earlier, a Nimrod topology map can be generated using as much aggregation as is required. Illustrated
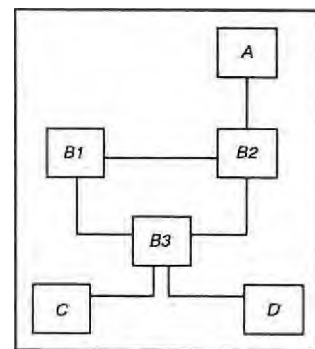
Figure 70 : Nimrod topology map – high level of abstraction.

in **Figure 70** and Figure 71 are two examples of Nimrod topology maps which represent the same physical topology shown in Figure 69. In the case of **Figure 70** the level of aggregation is high and their is little visible information of what actually happens within the nodes, particularly nodes B3, C and D. In **Figure 71** these nodes have been represented using less aggregation to illustrate the flexibility of Nimrod aggregation schemes.
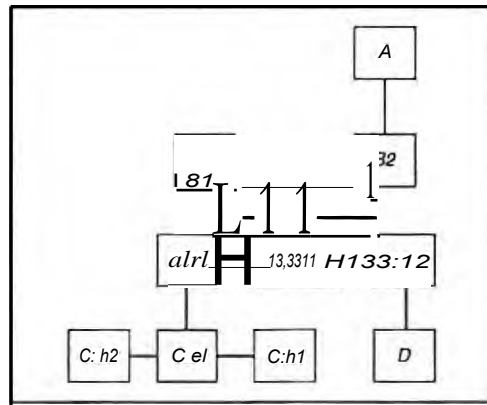
Figure 71: Nimrod topology map - variable level of abstraction.

It is important to notice that in all Nimrod nodes the internal map attribute can be used to determine more accurately what the internal structure of the network looks like, if this is not already explicit in the Nimrod topology map. In this example that would mean that the internal map attributes for node B3, C and D in **Figure 70** could contain the information which is made explicit for these nodes in **Figure 71.**

7.8.4 Cost and Benefit Analysis for Nimrod

After evaluating Nimrod against the criteria on chapter 6, some of the advantages and disadvantages which this protocol has to offer are discussed in this section. The aspects discussed are concerned particularly with addressing and routing issues, as were laid out in chapter 3.

Since Nimrod is a completely new subsystem[59] in the Internet architecture the cost of implementing the protocol is very high. However, in this light one must consider that the cost of implementing any change, be it large or small would still be very high [Clarke *et al,* 91]. Being able to implement a complete protocol subsystem can have its cost of implementation offset by the advantages which it brings. In the case of Nimrod an efficient and very scalable routing architecture are two of the most important benefits. Furthermore, the completely scalable address structure and the support for policy routing and Type Of Service requirements within the network are also very important advantages. Although some might say that not providing a mechanism for abstraction or route calculation within Nimrod was a serious shortcoming, it is in fact exactly the opposite. These two areas are ones where considerable change can occur as new techniques are developed, and being able to implement these changes without changing the base protocol makes Nimrod very flexible and powerful.

The choice of using a modified link-state algorithm to represent the network for Nimrod is a further benefit. This is the only feasible method for representing such a large scale internetwork and still be able to attach attributes to links without causing a huge state space explosion problem (See section 6.3) [Estrin *et al,* 93][Breslau-Estrin, 91].

An issue relating to flows within the intemetwork is that of having state information contained within the routers. In Nimrod the flow setup information is installed in the routers explicitly, and the datagrams then use a flow identifier for further routing. This decision is a robustness and efficiency trade-off, since a failure in any one of these involved routers would disrupt the flow (i.e. loss of state and the recovery from this position). The alternative of having each datagram contain all the flow setup information was deemed too expensive and inefficient to be acceptable. Other issues which supported the decision were the likely growth in the amount of user state required and the necessity for more reliable installation in routers.

Once a datagram is associated with a flow the forwarding process is purely hardware forwarding,

---

[59] The Nimrod subsystem - Nimrod provides the functionality for the Internet protocol in the following areas:; Routing information distribution, Route selection and User traffic handling.

which makes the process very efficient. In the case of the Nimrod datagram forwarding mode, this makes the forwarding task even more efficient than the current scheme in IPv4 which requires software forwarding (i.e has to perform a lookup for the next hop) [Chiappa, 94b].

Source routing traffic using either CSS, CSC or Flow mode forwarding has the advantage that it allows route policy to be determined by the traffic initiator. However it can be very costly in terms of processing. A less expensive route generation and forwarding mechanism such as suggested in [Estrin *et al,* 93], namely NR (Node Routing), has been implemented in Nimrod as the datagram forwarding mode to cater for traffic which does not require special policy considerations and is by its nature datagram traffic. This is a very important aspect for any **IP** protocol as a very large proportion of the IP traffic is still going to be datagram traffic (See section 2.3). This traffic must be catered for and should preferably be catered for in such a way as to be either equally efficient as the current hop-by-hop forwarding mechanisms or preferably even more efficient. Using the more complicated source routing forwarding techniques for simple datagram traffic would be costly and incur a high overhead, due to the relatively large amount of processing which has to be performed for a relatively small amount of data. (i.e. each packet source routed !)

Considering Nimrod's approach of using an address which is topologically significant and then some type of **EID** to identify the host, it implicitly provides support of mobile host and other multi-homing devices (See section 6.2.2) [Chiappa 2].

7,8.5 Summary of the investigation into Nimrod

Nimrod can be seen as offering a number of very important advantages over the existing IP protocol, which can be summarised as follows:

> Datagram traffic will continue to be supported using a very efficient forwarding mechanism. For the other three forwarding mechanisms the routes will be either partially or completely determined and specified by the initiator of the traffic, making them very efficient and allowing support for Type Of Service traffic. A new kind of address called a locator is created for use in routing decisions, although the old IPv4 address will be

retained, but as part of the **EID.** This solves the address overloading problem with the existing IPv4 address. Short term inter-operation with IPv4 is ensured as no host software has to be changed, although optimisations can be made to host software should this be required.

One cost which Nimrod has associated with it, is that a transition to a longer node identifier **or EID** is required for Nimrod to operate effectively across the Internet. This is going to be necessary anyway due to the growth of the Internet, and should not been seen as a major drawback of the protocol.

7.9 TP/IX

TP/IX is intended as a replacement protocol for the existing IPv4 network layer IP protocol. Implementing TP/IX involves making modifications to the TCP, UDP and ICMP transport layer protocols, as well as replacing the existing network layer protocol. For the purpose of this discussion we will only be considering the network layer protocol and any other aspects which directly influence addressing or routing within the Internet. The development philosophy behind TP/IX is *"Don't change things which work",* which makes it a somewhat conservative approach to solving the problems with IPv4.

7.9.1 What does TP/IX deal with?

The main areas of interest to us with the TP/IX protocols are the approaches taken to solving the addressing and routing problems associated with IPv4. The other areas which TP/IX deals with are the rewriting of the TCP and ICMP protocols to be able to make use of the new TP/IX address format.

❑ **Address space**

TP/IX increases the existing 1Pv4 address space from 32 bits to 64 bits. The modifications to the new address space include the addition of an Administrative Domain Identifier (AD)
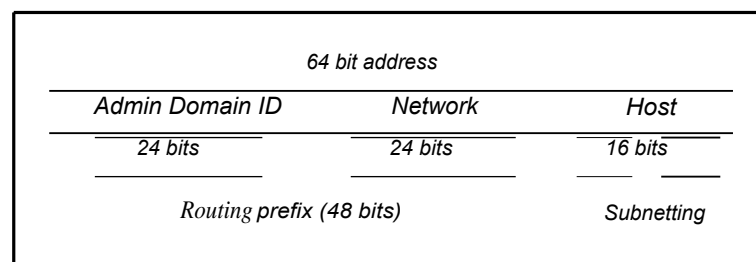
Figure 72 : Format for the TP/IX host address.

and the allocation of more bits within the address for subnetting purposes [Ullman, 93]. The format of the new address can be seen in **Figure 72.**

The appropriate changes have to be made to the DNS servers to indicate the new TP/IX address in its tables. To ensure minimal change it is critical that the AD identifiers are allocated from the IPv4 network numbers to prevent confusion of the top 4 octets of a **TP/IX** address with a IPv4 address. A typical zone entry within the DNS table of an updated system may look something like:

Garfield.I nterN IX.COM.   **A**   **192,24.59.86**

                              A   **192.0.0.192.24.59.1.86**

Where the Administrative Domain identifier for this example is 192.0.0 and the entry **Garfield.InterNIX.COM** has two **A**[64] field records in the DNS server [Ullman, 93]. For an interpretation of how the second **A** field entry is generated see **Figure 73.**

The definition of an administrative domain is a service provider or a national administration (government). The **TP/IX** address limits the number of ADs to 16 bits, which gives a theoretical limit on the number of ADs of 65536. At the moment there are only a few hundred ADs and growth is not expected to exceed a few tens of thousand ADs. This upper limit is the correct order of magnitude as the **number** of ADs have to be advertised at the Core routing" level and hence should not exceed the physical capabilities of the routing hardware.

The remainder of the address space is split between the network ID and the Host ID. The Host ID is allocated the lowest 16 bits in the address, which can then be subnetted even further if required by the particular network.

---

[64] A field - The A field within a DNS server contains the EP address for the host who's domain name is provided. [Stevens W.R., 94]

[6t] Core Routing - Routing done at the highest level within a hierarchical routing structure.

For communication between IPv4 hosts and TP/IX hosts the IPv4 address mapping into TP/IX address has been specified as shown in **Figure 73.** The first 3 octets of the address are the fixed AD identifier (which in this case has been allocated from the existing pool of IPv4 network numbers), followed by the first 3 octets of the IPv4 address. The remaining
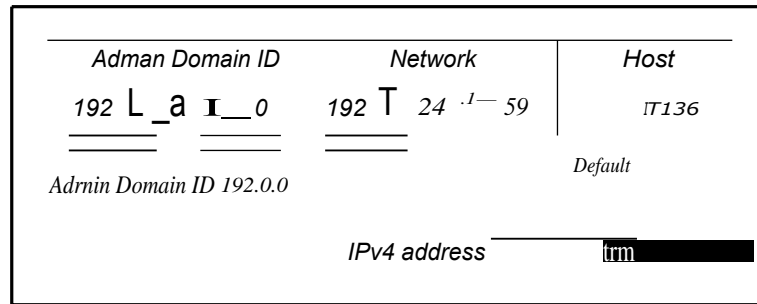
Figure 73 : IPv4 addresses mapping into TP/IX *addresses.*

2 octets of the **TP/IX** address contain the value 1_6 in the first octet and the remaining octet of the IPv4 address in the last octet.

Using this address mapping scheme an existing IPv4 address may be mapped into a TP/IX address as follows:

$$146.231.123.57 \text{ ip}_v\text{eF } 192.0.0.146231.128.1.57 \text{ }_{TP_A}x$$

where the AD identifier in this case is 192.0.0.

❑ **Routing** Issues

All routing within the TP/IX architecture is done using the top 48 bits of the address, that is the AD + Network ID portion of the address. The routing protocol to be used for TP/IX is equivalent in function to the existing CLNP[62] protocol, which provides similar

---

[62] CLNP - ConnectionLess-mode Network Layer Protocol is an ISO routing protocol with a similar functionality to the IP protocol.

service as IP, but with the advantage of being able to use bigger addresses [Piscitello D., 93].

ClIDatagram header optimisation

The datagram header for TP/IX has been optimised by excluding fields which are not

Version	Header length	Time to live (TTL)
	Total datagram length	
	Forward Route Identifier (FRO)	
	Destination address	
	Source address	
Protocol		Checksum
	Options	
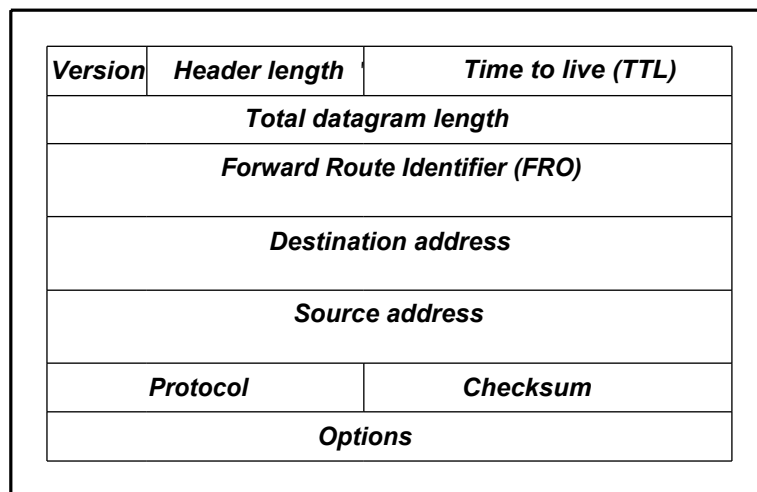
Figure 74 : TP/IX datagram header format.

absolutely necessary for all datagram traffic. At the same time the appropriate fields in the header have been increased in size to accommodate the larger addresses. Another change which has been implemented is the inclusion of a Forward Route Identifier (FRI) which can be used to optimise routing between consecutive routers in the network. The basic structure of a TP/IX datagram header can be seen in Figure 74. The version field (4 bits) indicates the current IP version namely version 7 (as incorrectly assigned by TP/IX). The header length field (12 bits) indicates the number of 32 bit words within the header. The TTL field is similar to the TTL field in IPv4, except that it is now a 16 bit field and each increment indicates 1/16 of a second. The total datagram length (32 bits) is indicated as the number of 32 bit words.

The FRI is a 64 bit field which contains a unique value for each datagram "flow[63]". This FRI field can be used to keep datagrams within reserved flows, perform mobile-host routing or a variety of other routing specific tasks. (See section 6.2.2) The protocol field (16 bits) indicates the transport layer protocol, e.g. TCP is 6.

Options may follow the basic header. These options have a variable length but must always be 32 bit aligned for processing purposes.

7.9.2 How does TP/IX work?

As there can effectively be two different addresses in the Internet when using TP/IX and IP, a number of scenarios for hosts operating using these different address types can occur. The standard case of two **IPv4** hosts exchanging traffic, using an existing IPv4 routing protocol is illustrated in **Figure 75** part A. The more interesting cases arise when either of the hosts use a different address type **or** the routing protocol does not support the host's address type. These cases will now be examined in more detail.

❑ **TP/IX and IPv4 host communication**

Upgraded or TPIIX hosts can communicate with IPv4 hosts, making use of the translation mechanism which is provided in the TP/IX hosts' code. The TP/IX host generates a native TP/IX datagram which is then translated into a IPv4 datagram before it is sent across the network using the existing IPv4 routing protocol **(see Figure 75** part B).

A number of different scenarios can result using these two host types. The other scenario which can occur is that of the routing protocol between the two hosts being an upgraded **CLNP** for TPIIX protocol. Even in this case the TPIIX host would have the responsibility

---

[63] "Flow" - The use of the word "flow" in this sense is used to indicate datagrams which are part of the same connection between a given source and destination(s).

of translating its datagrams, as the onus is always on the updated hosts to perform any datagram translation [Ullman, 93]. In this example the IPv4 host would have to be using an extended IPv4 address (exten&d to 64 bits), to be able to successfully communicate with the **TP/IX host.**

USame host types with an unmatched routing protocol.

In **Figure 75** part C shows two TP/IX hosts communicating using the existing IPv4 routing protocols. Since the hosts have 64 bit addresses which are incompatible with the existing IPv4 routing protocols, they have to use the IPv4 extended addressing option. This extended addressing option allows the IPv4 packets to carry a full length TP/IX address, and thus the hosts are still able to connect to other TP/IX hosts elsewhere in the network, even though the intermediate paths and routers may only be IPv4 capable.
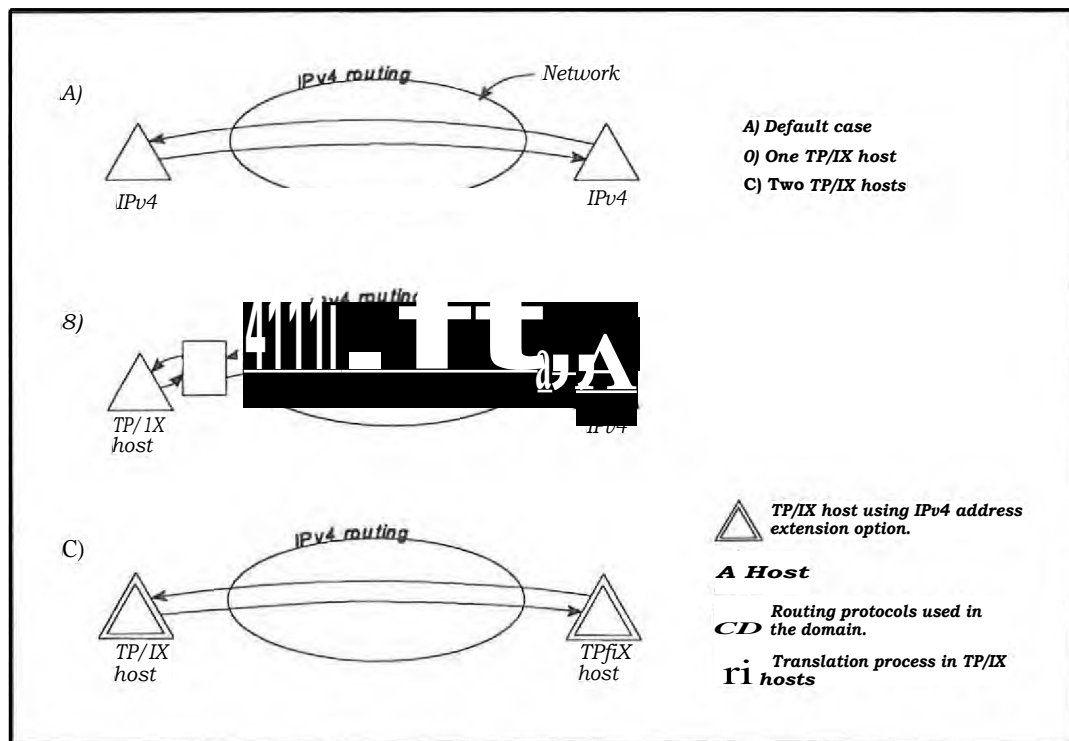
Figure 75 : Datagram routing and translation using TP/IX and IPv4 in the Internet.

The converse problem to this, where the two hosts are IPv4 and the routing protocol is TP/IX capable only, is solved in a similar fashion with both IPv4 hosts making use of the IPv4 address extension directly.

TP/IX hosts determine destination host types by first sending a DNS address request for the new extended **A** field (TP/IX address). If this request fails they then enquire the
. destination host's IP address and perform datagram translation on the datagrams.

The datagram translations which are performed by TP/IX hosts are effectively state-less, in an attempt to make the TP/IX and IPv4 host interaction as transparent as possible.

7.93 Cost and Benefit analysis of TP/IX

After analysing TP/IX against the criteria in chapter 6, the important aspects relating to this protocol, in particular addressing and routing will now be discussed.

TP/IX is designed to maintain the look and feel of IPv4, which can positively influence the migration from IPv4. (See section 6.1.1) Coupled with this familiar environment is the high degree of interaction provided between IPv4 and TP/IX hosts, using one of the techniques as described in the previous section.

The adoption of TP/IX is entirely user driven with as few transitional constraints as possible. Mechanisms for IPv4 to TP/IX translation (and reverse translation) are provided for the TP/IX hosts.

Requirements of the new IP protocol:

   Minimum number of supported networks 10^9
   Maximum number of entries in the core routing tables 20000-40000
   Very large number of hosts in the entertainment and device control markets.
   "Each subscriber may be thought of as an entire network supporting X hosts and not as a single host!" [Isted, 94]

Scaling properties of TP/IX addresses:

   Scaling of the AD (Administrative Domain Field)
         Hard wired to a maximum of 2^{16} ADs , but further limited by the physical router constraints. (Maximum size of the routing tables is typically 20K entries.)

         20000 entries in 24 bits 0.125% utilisation

   Scaling of the Network Field
         2^2 1.6 x 10^7 networks

   Scaling in the Hosts Field
         2^{16} 65000 hosts

Figure 76 : The scaling properties of TP/IX addresses.

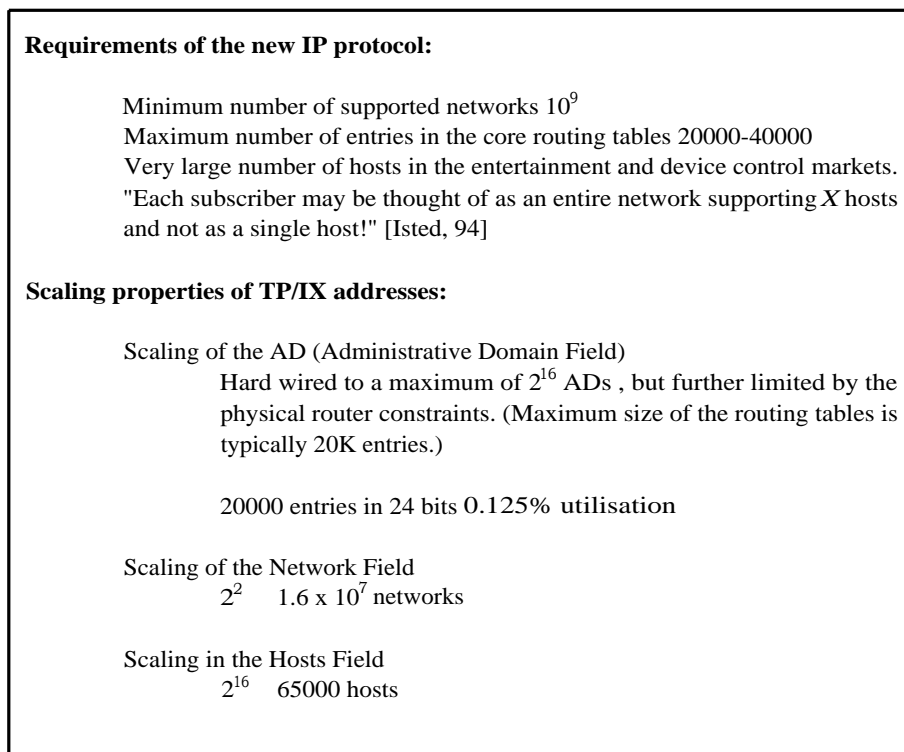The routing hierarchy as proposed in TP/IX does not provide sufficient aggregation to cope with very large internetworks. The aggregation which can be obtained from the TP/IX addresses is limited to three levels, namely the host level, network level and AD level (See section 6.3.2). Although subnetting is permitted within the host field of the address, the resulting aggregation is effectively invisible at any higher levels within the hierarchy.

The number of bits allocated within the TP/IX address is not sufficient to comply with the requirements for IPng. As seen in **Figure 76,** TP/IX can not support enough networks for the foreseen growth in the Internet. Not only this, but there is no mechanism in place to provide additional address space when the assigned address space is depleted (See section 6.3).

7.9.4 Summary of the investigation into TP/IX

The approach taken by TP/IX is too conservative. Despite solving the existing problems with the Internet Protocol, TP/IX has not provided for enough flexibility in its routing or addressing schemes to cater for the tremendous growth which is predicted for the Internet [Lottor, 92][Clarke *et al,* 911

The routing mechanism used in TP/IX, namely hop-by-hop routing, is not adequate as a routing mechanism for datagrams which require special Type Of Service requirements or other routing policy considerations. Despite hop-by-hop routing being very efficient for simple datagram traffic, there is likely to be a very substantial growth in traffic which is not routed optimally using this mechanism (See section 6.2).

The cost of implementing TP/IX is high due to the extensive modifications to host software, routers and DNS servers as well as to protocols within the transport layer. The benefits which TP/IX brings (increased address space) are sufficient to partially solve one of the most serious existing problem with IPv4 but they do not provide the flexibility to cater for future developments or problems within the Internet (See section 6.2.2 and 6.2.3).

7.10 Extended IP

The Extended Internet Protocol (EIP) is a proposal which aims at solving the problem of address space depletion by using a new scheme for addressing and routing. EIP is also designed so as to maintain maximum backward compatibility with the existing IP hosts in the Internet. Furthermore, EIP recognises that 1P is the cornerstone of the existing Internet and that replacing IP with some new protocol will inevitably require huge changes to a large number of the aspects of the current Internet.

7.10.1 What does EIP deal with?

EIP has been specifically designed to solve the address space depletion problem. At the same time the issue of maintaining maximal inter-operation between the existing IP hosts and the new EIP hosts has also been stressed.

To solve the problem of address space, EIP has effectively pushed the entire 32 bit IP address into the local domain, at the same time removing any of its routing significance. As a replacement to the network portion of the IP address, EIP relies on a new inter-domain addressing and routing protocol. This inter-domain addressing and routing protocol is not specified by EIP, since it is flexible enough to cater with a variety of candidates. The most notable candidates for this inter-domain addressing and routing protocol are Nimrod ,NSAP and PIP.

---

[64]Some of the aspects of the Internet which could be affected by a completely new IP are: routing, hosts, ARP, RARP, ICMP, TCP, UDP, HP, DNS and routers.
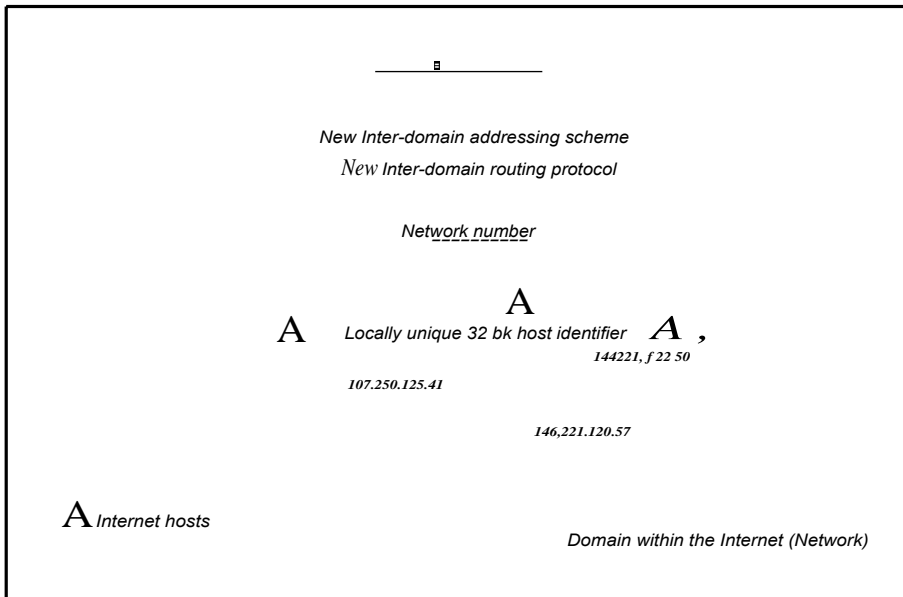
Figure 77 : The addressing scheme as proposed by FTP.

For an illustration of the changes which are involved to the addressing scheme refer to **Figure 77.** Notice that the host identifiers (no longer addresses as they are only locally unique) within a network are now taken from a 32 bit range. A completely specified address for a host is now specified as :

< routing header + host identifier >

With the exact format of the routing header determined by the specific inter-domain addressing and routing protocol implemented.

EIP maintains maximal inter-operability with **IP** by maintaining the existing **IP** datagram format
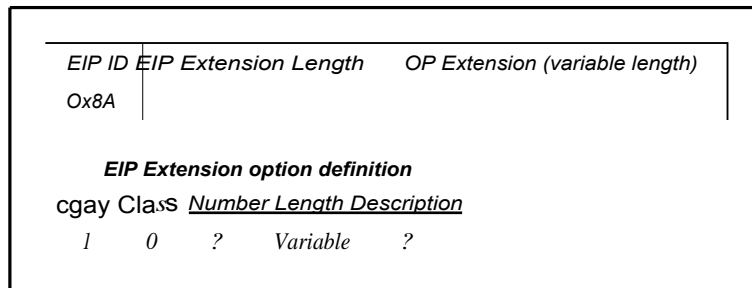
Figure 78 : Format of the FTP Extension option.

and IP address and merely implementing the EIP address as an option within the conventional IP protocol. The exact format of the EIP option as it is appended to the IP datagram can be seen in Figure 78 [Wang, 92b].

The EIP option is implemented as a variable length option, which means that the EIP Extension can allow for a variety of different inter-domain addressing mechanisms to be used. Stored within the EIP Extension is the source and destination network number, and any other fields which may be required by the inter-domain addressing and routing protocol. In the case of Nimrod these may be the source and destination End Point Identifiers [Chiappa, 934

7.10.2 How does EIP work?

For intra-domain communication the hosts operate exactly the same as with IPv4. Their host identifiers (old IP addresses) still have local uniqueness which means that the events local to the domain such as ARP, RARP, local DNS lookups and subnet routing remain exactly the same as with IPv4.
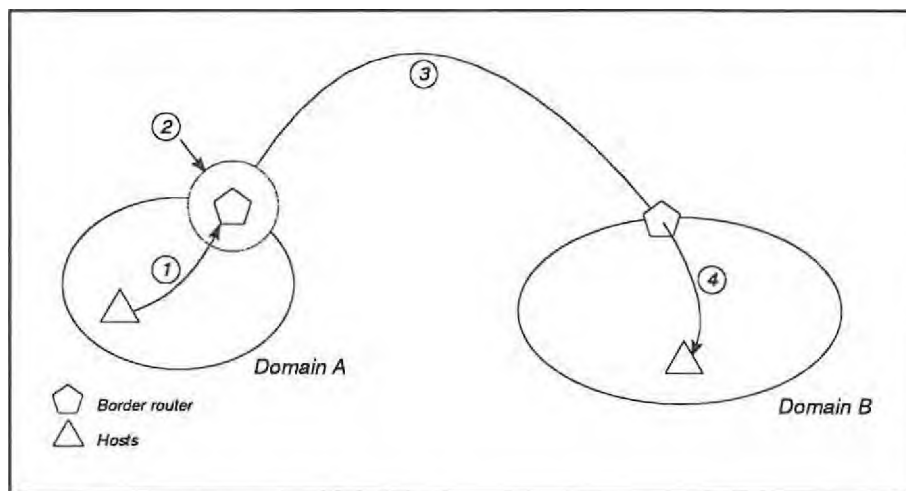
Figure 79   Inter-domain traffic using EIP.

In the case of inter-domain traffic, the EIP address becomes relevant. During the transitional stage when **32 bit IP addresses are still globally** unique the **scheme works as follows** (see **Figure 79) :**

9 The source host generates a datagram and sends it to one of its border routers.

CD At this border router the destination address (IPv4) is examined and the network portion removed and used in a lookup table to determine the destination E1P network number. The EIP source network number is also determined here, These EIP source and destination network numbers are then added into the EIP Extension of the datagram for future use.

CI At the inter-domain routing level the datagram is now routed according to the EIP addresses in the EIP extension by the new inter-domain routing protocol.

OD Once the datagram reaches the destination domain (network) it is then routed by the intra-domain routing protocol (IPv4 routing protocol) on the old IP address.

The case of using EIP once the 32 bit IP addresses are no longer globally unique is slightly different, in that the network portion of the IPv4 address can no longer be assumed to map to a unique EIP network number. In this case each host has to be able to use the entire EIP address to be able to communicate outside its own domain. Hosts which remain IPv4 compliant only will thus be limited to local communication [Wang, 92b].

7.10.3 Cost and Benefit Analysis of EIP

EIP has a very significant advantage over many other proposals in that there are no intra-domain changes necessary during the transitional phase. This in effect means that the hosts within the domains are unaware of the changes which are occurring outside their domains. It may have the negative effect of not promoting migration to upgraded hosts using full EIP addressing due to the notion that, *"everything still works, so why change?".* Users in domains who adopted this attitude could soon find themselves isolated and only able to communicate intra-domain, until they upgraded to fully EIP compliant software. (See section 6.1)

The changes which have to be implemented in intra-domain entities during the transitional phase are the upgrading of the DNS servers to include a new resource record to store the appropriate EIP network numbers. The new resource record has been denoted to be a type "N[65]" record while the host identifier then retains the type "A'" resource record.

An area where there may be a substantial cost associated, is at the border routers, where lookup tables containing IP network EIP network numbers have to be maintained. As this is only required during the transitional phase it may be tolerated, although any kind of lookup in border routers is highly undesirable as it causes a bottleneck for inter-domain traffic. Furthermore, the extensive processing of native IPv4 datagrams at border routers to include the EIP Extension option is another aspect which is not desirable, and can only be tolerated during the transitional phase.

Removing the transitional mechanisms after depletion of the IP address space may be seen as a serious problem by some, but it is believed that the added expense of running the translational services at border routers can not be tolerated for longer than the transitional phase. Apart from this issue, the proposed transitional mechanisms are not guaranteed to work once 32 bits address space depletion occurs. (See section 6.2.3)

EIP allows for the choice of a separate inter-domain addressing and routing protocol, which gives it a great deal of flexibility. This flexibility comes from the fact that the EIP protocol can focus on solving its particular problems and leave the separate issue of inter-domain addressing and routing to the appropriate protocol.

The decision to remove the routing significance from IPv4 addresses for the implementation of EIP is a very important. This break solves the problem with IPv4 addresses which were functionally over-loaded. As was pointed out in chapter 3, IPv4 addresses perform routing

---

[65] The type N resource record is a new type which will have to be created within the DNS servers code.

[66] Type A resource records within the DNS servers previously indicated the complete IPv4 address. Using EIP they now indicate the host identifier ,which has no global routing significance.

functions as well as host identification, With the adoption of EIP they are no longer addresses but merely 32 bit host identifiers which have no other significance. Using the scheme of a 32 bit host identifier which has to be unique within its own domain only allows for 4×10^9 hosts per network. This is more than enough to cater for the foreseeable future. (see sections 6.3.1. and 6.3.2).

The scaling properties of the number of networks in the Internet is determined by the inter-domain addressing and routing protocol.

EIP upgrading is fairly flexible within domains until the end of the transitional stage, after which the upgrade plan is fairly strict. This should not be a major problem as the time available to upgrade within domains is relatively long (the entire length of the transitional phase). Obviously the first major phase of the protocol upgrade affects the inter-domain entities such as the border routers, backbone routers and some DNS servers. Other modifications which have to be made to the local domains entities before the end of the transitional phase are:

- Modification of FTP to exclude the IP address from the Data Port command.

- Modification of local DNS servers to include the N resource record

- Modification to key hosts such as mail sewers, FTP servers and central machines to support EIP.

- Modification to subnet routers to support EIP routing

7.10.4 Summary of the investigation into EIP

Although EIP does not solve the routing problem, it solves the address space problem very well and allows for the routing problem to be solved externally to its problem domain. Implementation costs for EIP are relatively low, especially during the transitional stage during which the translation mechanisms take care of most of the EIP to IP translation.

EIP certainly succeeds well in the area of maintaining maximal IP compatibility, which is of great importance if it is going to succeed as an IP replacement.

7.11 Conclusion

The 10 protocols which have been examined bring to the fore a multitude of different approaches to solving the same problems. Not all of the examined protocols are as appropriate for use in the Internet due to some of their shortcomings, however the merits of each has to be acknowledged and could perhaps be incorporated into the final protocol which will replace IP in the future.

The next chapter examines the overall performance of these protocols and makes some suggestions as to what properties the next IP candidate should posses.

8. Findings and recommendations

8.1 Summary of Findings

The findings as reported in chapter 7 can be summarized in **Figure 81** on page 167. **Figure 80** (on page 166) refers specifically to the findings relevant to the problems listed in chapter 3.

Firstly, the proposals are all classified as to their suitability for long term or short term implementation, see column A and **B** of **Figure 80.** Short term implementation means that the proposal is sufficient to solve some of the existing problems, but not sufficient to cater for the future requirements of the Internet (see chapter 6). Some protocols such as CIDR and CATNIP have the property that they could be implemented as both long or short term protocols.

The address depletion and routing scaling problems (see chapter 3), are the primary interests of this thesis. Consequently the proposals are further classified according to whether they solve one or both of these problems. Partial solutions are proposals which only solve one problem successfully, whereas complete solutions solve both problems.

    CI Partial solutions :

    A number of the proposals only solve a subset of the problems presented in chapter 3 (see column C in **Figure 80).** These protocols (CIDR, AEIOU, NAT, Two-Tiered, EIP) should not be completely excluded since they can often be combined with other protocols which solve the remaining problems. (e.g. NAT used with TUBA, where NAT is used for the inter-domain traffic and TUBA solves the inter-domain addressing problem.)

    CIDR is a partial solution as it only solves the routing problem, and does not examine the address depletion problem. Similarly AEIOU, NAT, Two Tiered, and EIP are also partial solutions as they only address the address space depletion problem and provide no

solutions to the routing problem. Furthermore, these solutions are partial solutions as they can only provide a temporary solution to the address space depletion problem, at the cost of global connectivity (Hosts in stub domains which implement these protocols generally loose the global uniqueness of their addresses).

❑ **Complete solutions :**

Four of the evaluated protocols are complete protocols and could be implemented as a replacement for **IP.** These protocols are: TUBA, **SIPP 16,** Nimrod, and TP/IX all provide for a larger address space. Neither SIPP 16 or TP/IX place much emphasis on the Internet routing problems (See chapter 3). Since the routing problem is one of the more serious problems with the existing Internet, any contender for the next **IP** protocol will have to address this problem more earnestly. Nimrod and TUBA present proposals which have very good routing scaling properties and at the same time provide a very large address space.

The proposal presented by CATNIP does not fall into the class of either partial or complete solution. The CATNIP proposal has a completely different function, namely to provide a common Network Layer, through the use of a Common network layer address and Common Network Layer Datagram. Using these mechanisms CATNIP provides inter-protocol communication without requiring translation mechanisms in each participating protocol.

8.2 Recommendations

The author believes recommendations can be made for several distinct areas within the IP protocol. These recommendations do not necessarily correspond to the recommendations of any one protocol, but are a summarized synthesis of ideas sourced not only from the protocols already put forward in the literature, but from discussions which the author has held with other parties as well.

❏ **Addressing :**

The over-loading of the term address has to be solved by using a non-topologically related **EID** (Endpoint Identifier). At the same time a clear distinction must be made as to which information is used for routing and which for flow selection or Type Of Service considerations. Refer to column D in **Figure 80** for the various protocols address formats and address space. According to the above criteria, the layout for a completely specified address should be as follows:

routing information + selector + destination identifier

None of the proposals define addresses in this way. Nimrod recognizes the importance of the separate routing information and the destination identifier, but lacks the selector field. The selector field is important for the processing of addresses by routers. Using the selector field in the address sequence, the router can efficiently determine if the datagram requires specialized handling. The routing information must be topologically significant, whether this means a series of domain identifiers or some hierarchical **address, is** determined by the network representation method (Multi-level hierarchy or unstructured topology).

The next criteria for host addressing is that it be large enough to cater for the expected future growth (see chapter 2). This places the lower limit on the number of bits which can be used for host addresses at 32 bits. Referring to column D in **Figure 80,** we see that most of the protocols have sufficient address space, provided that they do not use additional bits within this space to create some hierarchical structuring (this would decrease their available address space for host addresses). A further requirement for addresses, is that of providing for enough networks within the address space. Network address space, as shown in column E of **Figure 80,** should be capable of catering for all future network growth in the Internet. An added requirement on the network addresses, is that they aggregate sufficiently to enable efficient routing and small routing **tables. This**

efficient aggregation is strongly coupled with the hierarchy within the network addresses. See column F **in Figure 80** for the hierarchy information of the complete protocols.

The most suitable address of all the protocols examined, is that of the Nimrod protocol, even though it does not match the ideal address format as suggested above. This address is a variable length address with variable levels of hierarchy, which enables it to cater for any eventual network growth.

❑ **Routing :**

Routing is the issue which is causing the severe problems with the existing Internet. With the predicted growth in the Internet there is a very serious need for the abstraction of routing information[67] in the Internet. Since information abstraction is only possible with some form of structured topology, this places a constraint on the type of network representation which can be chosen (See column F and G in **Figure 80).**

The ideal topology representation would be one similar to that used by **Nimrod, which is flexible in** the sense that it is hierarchical but has no predefined "top" or "bottom" **level. Within** this hierarchy the topology can continue to be expanded without damage to the abstraction mechanism or placing a limit on the growth of the network size or complexity.

❑ **Network representation" :**

The network representation (see column G **in Figure 80)** is a crucial factor determining

---

[67] Abstraction of routing information - Routing information is that information which is passed between routers to enable them to build a network topology map. Abstraction of this information implies that less information needs to be exchanged between routers. Despite the abstraction there are no adverse affects on the completeness of the topology map.

[63] Networkrepientation - The particular representation technique used to represent the network, its connectivity and any other attributes which it may have.

the routing protocol algorithm and types of advanced routing services which can be offered (see column H in **Figure 80** for the routing mechanisms which are supported by the various protocols). The initial IP specification defined the use of a distance-vector representation. This representation works very well for datagram traffic, but is hopelessly inefficient for more advanced routing services (e.g. policy routing and other Type Of Service considerations).

The author's choice for a replacement for the distance-vector representation is that of the modified multi-level link-state representation as introduced by the Nimrod protocol. This selection allows better support for more advanced routing services and more convenient representation techniques. Coupled with an efficient abstraction mechanism the multi-level link-state representation also scales very well. The scaling criteria are crucial, since this is one area where the distance-vector representation falls short once link attributes are included. (Link attributes are needed to satisfy Type Of Service and flow requirements.)

❑ **Protocol migration issues :**

With the adoption of any new protocol there are the associated migration issues, backward compatibility problems and upgrade dependancies. **Figure 81** presents a summary of these effects for each protocol.

Migration to any new protocol which uses a different addressing technique is going to affect network entities such as DNS servers and border routers (See the Migration column in **Figure 81).** Since every protocol examined requires significant changes to the existing IPv4 network entities in this area, it would be well worth the effort implementing a protocol such as CATNIP. CATNIP not only eases inter-protocol operation during protocol transition but also provides for continued inter-protocol communication between numerous protocols even after the transitional stage.

CATNIP can be used to provide extensive compatibility between IPv4 and the new IP

protocol (see the middle column in Figure 81). CATNIP is not restricted to providing compatibility between only two protocols, but can be used for multi-protocol communication. The existing CATNIP specifications already allow for interaction between IPv4, SIPP, CATNIP, and NSAP datagrams.

Finally, some of the upgrade dependencies for each of the protocols are listed in the last column of **Figure 81.** Typically, the more complete (i.e. solves more problems) the protocol, the more complicated the upgrade is going to be and hence the likelihood for more upgrade dependencies to develop.

Figure 80 :Findings related to addressing and routing issues

Protocol	A	C	Solution to ..	Addressing Space for hosts	Addressing Space for networks	Routing Hierarchy	Network representation method	Routing mechanism
CIDR	X	x	Address Allocation & routing costs	Dependant on complete address size (Bitwise maskable)	Dependant on complete address size (Bitwise maskable)	Dependant on routing protocol. Murti-level if CIDR capabte.		It I al capabte routing protocot, for bit maskabre forwardingg.
AE1OU		x	Address depletion	32 bits for AE (Host identifier)	32 bits for SI (Network prefix)			
NAT		x	Address depletion	32 bits locally significant addresses	Same as before NAT implementation			
Two-Tiered		x	Address depletion	subset of 32 bits for temporary addresses (globally unique) and 32 bits for permanent addresses	32 Bit globally unique network numbers			
TUBA	x		Address Depletion and Routing	NSAP addresses as for the GOSIP ver 2 format 56 bits possible + hierarchy	NSAP addresses as for the GOSIP ver 2 format 56 bits possible + hierarchy	NSAP murti-lever routing hierarchy	Link State representation for CLNP	CLNP (ConnectionLess Network Protocol)
SIPP 16	•		Address Depletion and Routing	128 bit addresses Host portion dependant on the address type and suboetting	128 bit addresses. Dependant on the address type.	Multi-level hierarchy	Link State representation for an upgraded CIDR routing protocol.	OSPF or BGP-4 upgraded to *use* the larger addresses.
CATNIP	1	x	Address Depletion and Routing	Dependant on Transport Layer protocol.	Dependant on Transport Layer protocor.	Dependant on Transport *Layer* protocol.	Dependant on Transport Layer protocol.	Dependant on Transport Layer protocol.
Nimrod	a		Address Depletion and Routing	Variable length : Multi lever rocators	Variable length ElDs	Multi level hierarchy using Nimrod areas (abstraction)	Multi - level Link State map with abstraction techniques	CSS, CSC, DM, FM (Sec section 7.8.2)
TP/I X	X	x	Address Depletion and Routing	64 Bit address with 16 Bits for host identifiers	64 Bit address with 24 bits for networks and a rimit of 40000 on the number of ADs	Multi level using the AD and Network portion of the address,	Dependant on the routing protocol	Hop by hop routing / Dependant on the routing protocol
EIP	x		Address Depretion and Routing	32 bit 1?v4 address which is localry unique	32 bit El? portion which is gtobalry unique			

Protocol	Migration (Before and After)	Backward Compatibility	Changeover requirements and dependancies
CIDR	Affects routing protocors and address assignment mechanisms.	CIDR aggregate information exptoded at domain boundaries, where. CIDR is not supported.	CIDR capabre intro-domain muting protocols required in the stub domains.
AEIOU	Affects muting software and client software	Only compatibility between partly specified hosts and IPv4 hosts or partly specified hosts and compretery specified hosts.	Host and muter software upgrade required
NAT	Affects DNS servers and border routers	FTP, NIPS and other apprications using LP addresses internarly will fail.	Border routers require new software and requires the installation of NAT hoses in the stub domains.
Two-Tiered	Affects DNS servers and border roisters	FTP, NFS and other applications using IP addresses internally wirl fail	DNS servers require a new resource record border routers mute on the 32 bit gtobally unique address. PASS (External Address Sharing Server) servers required.
TUBA	Iı˚ addresses migrate to NSAP addresses IP routing migrates to CLNP routing	Datagram encapsulation is used, although applications making explicit use of LPv4 addresses wilr riot work with TUBA.	Incremental deployment whilst IPv4 is operating until the entire stub domain (DNS, hosts, routers) is TUBA compliant and the relevant software has been updated.
SIPP 16	IP addresses migrate to SIPP 16 addresses IP routing migrates to SIP? 16 muting (extended CIDR type routing protocols)	Address encapsulation is ˙tccel, although applications making explicit use of IPv4 addresses will not work with SIPP, these applications have to be modified.	IPv4 and SIPP 16 coexist and interact to a smarl extent A complete software upgrade for arl network entities is required
CATNIP	Transport rayer address ,... CATNIP address Transport Layer datagram ••• CATNIP datagram	Very good inter-protocol compatibirity due to the Common network layer address format and Common network Layer datagram format	Incrementally deproyable network layer with no dependancies, Migration to the CATNIP network layer is invisible to the **users.**
Nimrod	IP address becomes a Nimrod locator and HD Routing mechanisms also affected to use the appropriate link slate muting algorithmc.	Initially IPv4 versions of FTP, NFS and other applications using IP addresses internalry may wort if the IPv4 address has been mapped directry into the corresponding EID. These applications should be rewritten to exclude the explicit use of the IP address.	Completely new protocol which requires extensive software upgrades, will provide limited IPv4 and Nimrod interaction,
TP/IX	IP muting becomes modified CLNP muting IP address becomes a 64 bit TP/IX address New TP/IX network layer. **Ma** intains a IPv4 rook and feel	FTP, NFS and other applications using IP addresses internally will fail..	User driven changeover. Modifications required to DNS servers (type A resource **record** needed) TCP, IMP and ICMP software.
EIP	IPv4 address becomes EIP host identifier IPv4 datagram = EIP datagram Variety of muting protocols accepted. (Nimrod NSAP PIP)	Very good backward compatibility until IPv4 address depletion at which point all translation mechanisms arc removed.	User driven changeover until IPv4 address depletion at which point there will be very limited IPv4 interaction and a strict upgrade path (FTP, DNS, servers, subset routers). Modifications required to DNS servers (type N resource record required)

83 Future Work

This thesis has looked almost entirely at the addressing and routing implications of the existing IPng proposals. There are numerous other related fields of research which still need to be investigated.

83.1 Transitional mechanisms

Much needed research is required in the field of transitional mechanisms between IPv4 and IPng. Transitional mechanisms deal extensively with the conversion between IPv4 addresses and datagrams and IPng addresses and datagrams. Issues which have to be examined are:

- Converting datagram header options which do not correspond directly between IPv4 and IPng.

- Datagram handling policies, when the different protocols define different handling policies (i.e maximum datagram sizes and fragmentation policies).

- Address encapsulation techniques are required which still uniquely distinguish a host within the Internet. These techniques have to consider the case of complete 32 bit IPv4 address depletion, when no new 32 bit addresses will be globally unique.

- The most efficient positions have to be determined for the various transitional mechanisms within the Internet.

- Modifications required by these transitional mechanisms to other network entities such as DNS servers and routers have to be investigated.

The goal of any transitional mechanism should be to maximize inter-operability whilst retaining minimal cost.

8.3.2 Routing mechanisms for use with link state network representations

A further field which is in need of much research is that of routing mechanisms for use with link-state network representations. The mechanisms as used for distance-vector representations are not directly transferable and issues such as routing in dynamically changing topologies and routing for Type Of Service requirements still need much attention.

Other areas of interest are those relating to determining the minimal amount on state information which has to be stored for each link to satisfy the service requirements of IPng. The efficient processing of this state information and well as its dissemination through the Internet are further areas which require attention.

8.33 Data security and encryption

With the increased commercialization of the Internet there is an increased requirement for adequate data encryption techniques and data security. If IPng is to be a commercially viable protocol, it has to offer levels of data security which are acceptable for business use, Research relating to how these two requirements impact on the overall design of IPng is required.

8.3.4 More efficient network management tools

The existing problems with the Internet are predominantly related to physical limitations of the IP protocol and hardware. Another alarming trend which is becoming increasingly visible is that the management of the state information required to keep the Internet operating is fast exceeding human capabilities. Already vast amounts of information have to be manually updated in routing tables, and many network entities (e.g. hosts running as special servers) require complex human configuration. Research is urgently required into the automation of these complex configuration and maintenance tasks, before the Internet reaches the point where it cannot be managed due to human limitations.

9. Conclusion

This thesis examined a number of the proposals which have been submitted as replacements for the IP protocol. It concentrated on addressing and routing related issues, as these are the two most serious problems facing IP. The addressing problem was shown to be related to the depletion of certain address classes and to the overall depletion of IP address space, whilst the routing problem was primarily concerned with the growth of routing tables and the ineffective aggregation of routing information.

Different protocols were evaluated against a set of criteria which were identified as necessary for the next generation IP. These criteria were based on growth patterns of the Internet as well as other work done in the area and included the scaling abilities of both host addresses and network addresses, routing mechanisms, and network representation techniques.

It was found that many of the proposals placed more emphasis on solving the address space depletion problem, than the more serious routing problem. The proposals examined presented both long and short term solutions, of which CIDR, Nimrod, and CATNIP were the most realistic, catering for the future growth of the Internet. CIDR (which has already been implemented at some inter-domain levels) was found to provide a good basis for the aggregation of routing information. Nimrod was found to contain the most flexible and unrestricted addressing mechanism, as well as a very suitable network representation technique for large internetworks. CATNIP was found to provide a generic platform for protocol inter-connection and hence backward compatibility. These three proposals could form the basis of a new IP protocol which would not face the same growth problems as the existing IP protocol, and at the same time would cater for a wide variety of future applications and network requirements.

References

[Aziz, 93] Aziz, A.,"A *Scalable and efficient intra-domain tunnelling mobile-IP scheme",* ACM Sigcomm, Vol 23, 1993

[Bellovin, 94] Bellovin, S., *"Security Concerns for IPNG",* RFC 1675, August 1994

[Braden, 89] Braden , R. , *"Requirements for Internet hosts - communication layers",* RFC 1122, October 1989

[Breslau-Estrin, 91] Breslau, L. ,Estrin, **D.,** *"Design and Evaluation of Inter-Domain Policy Routing Protocols" ,* ACM Sigcomm, Vol 20, 1990.

[Callon, 92] Callon , R. *,"TUBA: A proposal for addressing and routing",* **RFC** 1347, June 1992

[Chiappa, 94a] Chiappa, N., *"IPNG technical requirements of the NIMROD routing and addressing architecture",* RFC 1753, December 1994

[Claffy *et al,* 94] Claffy, K.C., Braun H.W., Polyzos, G.C., *"Tracking long term growth of the NFSNET",* CACM, August 1994

[Claffy *et al,* 93] Claffy, K., Polyzos, G., Braun, H., *"Traffic characteristics of the T1 NSFNET backbone",* ACM Sigcomm 93, Vol 23, 1993

[Clark *et al.,* 91] Clark, D., Chapin, L., Cerf, V., Braden, R., Hobby R.," *Towards the Future Internet Architecture",* Network Working Group, RFC 1287, December 1991

[Clark, 88] Clark, D., *"The design philosophy of the DARPA Internet protocols",* ACM SIGCOMM , Vol 18, No 4, August 1988

[Collela, 91] Collela R., Gardner E., Callon *R.,"Guidelines for OSI NSAP Allocation in the Internee',* RFC 1237, July 1991

[Corner, 91] Corner D., *"Internetworking with TCP/IP : Principles, Protocol, and Architecture Second edition",* Prentice-Hall International, 1991

[Deering, 88] Deering, S., *"Multicast routing in internetworks and extended LANs",* ACM SIGCOMM, Vol 18, No 4, August 1988

[Deering, 92] Deering, *S.,"SIP protocol specification",* working draft, 1992

[Dixon, 93] Dixon, *T.,"Comparison of Proposals for the Next version of IP",* RFC 1454, May 1993

[Droms, 93] Droms, *R.,"Dynamic Host Configuration Protocol",* RFC 1541, October 1993

[Egevang *et al,* 94] Egevang, K., Francis, P.," *The IP Network Address Translator (NAT)",* RFC 1631, May 1994

[Eidnes, 94] Eidnes, H., *"Practical Considerations for networking addressing using CIDR",* CACM, Vol 37 no 8, August 1994

[Eriksen, 94] Eriksen, H., *"MBONE: The Multicast Backbone",* CACM, Vol 37 no 8, August 1994

[Estrin *et al,* 94] Estrin, 11, Li, T., Rekhter, *Y .,"Unified Routing Requirements for IPng",* RFC 1668, August 1994

[Francis *et al,* 94b] Francis, P., Govindan, R., *"Flexible Routing and Addressing for a next generation IP",* ACM Sigcomm, Vol 24, No 4, October 1994

[Francis, 94a] Francis, P., *"Pip Header Processing",* RFC 1622, May 1994

[Francis, 94c] Francis *P.,"Pip Near-term Architecture",* RFC 1621, May 1994

[Fuller *et at,* 93] Fuller , V. , Li. , T. , Yu. , J. , Varadhan. , K. , *"Classless Inter-domain Routing (CIDR) : An address assignment and aggregation strategy",* RFC 1519, September 1993

[Fuller *et al,* 92] Fuller V. Li. T Yu J., Varadhan *K.,"Supernetting: An address assignment and aggregation strategy",* RFC 1338 , March 1992

[Gerich, 93] Gerich, E. *,"Guidelines for the management of IP address space",* RFC 1466, May 1993

[Goodman *et al,* 94] Goodman S.E., Press, L.I., Ruth, S.R., Rutkowski, A.M., *"The Global Diffusion of the Internet: Patterns and Problems",* CACM, Vol 37 No 8, August 1994

[Gross *et al,* 92] Gross, P., Almquist, P., *"IESG Deliberations on Routing and Addressing",* RFC 1380, November 1992

[Hemrick, 85] Hemrick C., *"The OSI Network Layer Addressing Scheme, Its implications, and Considerations for Implementation".* NTIA report 85-186, U.S. Department of commerce, National Telecommunications and Information Administration, 1985

[Hinden, 94b] Hinden, R., *"Simple Internet Protocol Plus White Paper",* RFC 1710, October 1994

[Hinden, 93] Hinden , R. *,"Applicability Statement for the implementation of Classless Inter-*

domain Routing (CIDR)", RFC 1517, September 1993

[Information Sciences Institute USC, 81a] Information Sciences Institute, University of Southern California, *"Transmission Control Protocol : DARPA Internet Programs Protocol Specification",* RFC 793, September 1981

[Information Sciences Institute USC, 81b] Information Sciences Institute, University of Southern California,*"Internet Protocol : DARPA Internet Programs Protocol Specification",* RFC 791, September 1981

[Ioannisdis, 94] Ioannisdis, J., Duchamp, D., Maguire Jr, G., *"IP-Based Protocols for Mobile Internetworking",* CACM, Vol 36, 1993

[Isted, 94a] Isted, E.D., *"Extendable addressing and its implications on Internet Protocol design",* S.A. conference for Computer Science Postgraduates

[Kahn, 94b] Kahn. R., *"The role of the government in the evolution of the Internet",* CACM, Vol 37, No 8, August 1994

[Kahn 94a] Kahn R.E.,*"Viewpoint-The role of government in the evolution of the Internet",* CACM, Vol 37, No 8, August 1994

[Kleinrock, 94] Kleinrock, L., *"Nomadic Computing - An opportunity",* ACM Sigcomm, Vol 25, August 1995

[Leiner, 94] Leiner, B., *"Internet Technology",* CACM, Vol 37 no 8, August 1994

[Leiner *et al,* 93] Leiner, B., Rekhter, Y., *"The Multiprotocol Internet",* RFC 1560, December 1993

[Lottor 92] Lottor M., SRI International," *Internet Growth (1981-1991)",* RFC 1296, Network Working Group, January 1992

[McGovern *et al,* 94] McGovern M., Ullman *R.L.,"CATNIP: Common architecture for the Internet",* RFC 1707, October 1994

[Piscitello, 93b] Piscitello D., *"Assignment of System Identifiers for TUBAICLNP Hosts",* RFC 1526, September 1993

[Piscitello, 93a] Piscitello *D.,"Use of ISO CLNP in TUBA Environments",* RFC 1561, December 1993

[Piscitello, 94] Piscitello, *D.,"FTP Operation over Big Address Records (FOOBAR)",* RFC 1639,

June 1994

[Postel, 81b] Postel, *J.,"Internet Control Message Protocol - DARPA Internet Program Protocol Specification",* RFC 792, September 1981

[Postel, 81a] Postel, J. ,*"Internet Protocol - DARPA Internet Program Protocol Specification",* RFC 791, September 1981

[Rekhter *et al,* 93a] Rekhter , Y. , Li. , T. , *"An architecture for IP address allocation with CIDR",* RFC 1518, September 1993

[Rekhter *et al,* 93b] Rekhter, Y., Topolcic *C.,"Exchange Routing Information Across Provider Boundaries in the CIDR environment",* RFC 1520, September 1993

[Saunders, 95] Saunders, S., *"Next Generation Routing - Making Sense of the Marketetures",* Data Communications, September 1995

[Simpson, 94] Simpson, *W.,"IPng mobility considerations",* RFC 1688, August 1994

[Stevens, 94] Stevens W.R., *"TCPIIP Illustrated Volume 1 - The protocols",* Addison Wesley, New York, 1994

[Symington *et al,* 94] Symington, S., Wood, D., Pullen, M., *"Modelling and Simulation Requirements for IPng",* RFC 1667, August 1994

[Taylor, 94] Taylor, M., *"The Cellular Industry view of IPng",* RFC 1674, August 1994

[Teraoka *et al,* 94] Teraoka, F., Yokota, Y., Tokoro, M.,"A *Network Architecture Providing Host Migration Transparency",* CACM ,Vol 37 no 8, August 1994

[Teraoka *et al,* 94] Teraoka, F., Uehara, K., Sunahara, **H.,** Murai, J., *"VIP: A Protocol Providing Host Mobility",* CACM, Vol 37 no 8, August 1994

[Topolcic, 93] Topolcic, C., *"Status of CIDR deployment in the Internet",* RFC 1467, August 1993

[Tsuchiya *et al,* 93] Tsuchiya, P., Eng, T., Computer Communications Review, *"Extending the IP Internet through address reuse",* ACM Sigcomm, Vol 23, No 1, January 1993

[Tsuchiya, 91b] Tsuchiya, *P.,"On the assignment of Subnet numbers",* RFC 1219, April 1991

[Ullman, 93] Ullman *R.,"TPIIX : The next Internet",* RFC 1475, June 1993

[Wang, 92b] Wang *Z.,"EIP : The Extended Internet Protocol - A framework for maintaining*

Backwards compatibility", RFC 1385, November 1992

[Wang *et al,* 92a] Wang, Z., Crowcroft, *J.,"A two-tiered Address Structure for the Internet : A solution to the problem of address space exhaustion.",* RFC 1335, May 1992

Other Sources:

[Carlson *et al,* 94] Carlson, R., Ficarella, D., *"Six virtual inches to the left: The problems with IPNG",* Memo of the networking group, May 1994

[Carpenter, 94] Carpenter, B., Work in progress, CERN, Network Working Group, 21 March 1994

[Castineyra *et al,* 94] Castineyra, I., Chiappa, N.J., Steenstrup, M., *"The Nimrod routing architecture",* Internet draft -work in progress, July 1994

[Castineyra *et al,* 95] Castineyra, **I.,** Chiappa, N.J., Steenstrup, M., *"The Nimrod routing architecture",* Internet draft -work in progress, March 1995

[Chiappa, 93b] Chiappa, *N., "Minutes of the New Internet Routing and Addressing Architecture BOF (NIMROD)",* Electronic document, November 93

[Chiappa, 95] Chiappa, *N.,"Nimrod architecture",* work in progress, March 1995

[Chiappa, 94b] Chiappa, *N., "The IP addressing Issue",* Internet draft - work in progress, April 94

[Chiappa, 93a] Chiappa, N., *"A new IP routing and Addressing architecture",* Electronic document, 1993 (Nimrod web page)

[Crocker, 95] Crocker D. Hinden R. *,"Simple Internet Protocol Plus (SIPP) - working group* charter",electronic document,

[Curran, 94] Curran, J., *"Market viability as a an IPng criteria",* Internet Draft-work in progress, March 1994

[Deering, 94b] Deering *S.,"Simple Internet Protocol Plus (SIPP) Specification ",* working draft, February 1994

[Deering *et al,* 94c] Deering S., Francis **P.,** Govindan **R.,** *"Simple Internet Protocol Plus (SIPP) : Routing and Addressing",* working draft, February 1994

[Deering, 94a] Deering S., *"Simple Internet Protocol Plus (SIPP) Specification (128 Bit address version)",* working draft, July 1994

[Droms, 92] Droms, R., *"Dynamic Host Configuration Protocol",* Work in progress, March 92

[Estrin *et al.* 93] Estrin, D., Rehkter, Y., Hotz, S., *"Scalable Inter-Domain Routing architecture",* Electronic document, 1993

[Francis *et al,* 94d] Francis P., Deering S., Hinden R., Govindan R., *"Simple Internet Protocol Plus (SIPP) : Addressing architecture",* working draft, July 1994

[Gilligan, 94] Gilligan R., Nordmark E., Hinden *R.,"IPAE : The SIPP Interoperability and Transitional mechanism",* working draft, March 1994

[Govindan *et al,* 94] Govindan, R., Deering, S., *"ICMP and IGMP for the Simple Internet Protocol Plus (SIPP)",* work in progress, March 1994

[Hinden, 95] Hinden, *R.,"IP Next Generation Overview",* electronic document, May 1995, IPng web site

[Hinden, 94c] Hinder, *R.,"On the design of an Internet Layer and the replacement of the Internet Protocol",* electronic document, 1994

[Hinden, 94a] Hinden *R.,"Simple Internet Protocol Plus White Paper",* working draft, February 1994

[IPng news group, 95] IPng news group (Big-Internet@Munnari) and archives during 1994 and 1995

[Isted, 94b] Isted, E.D, *"Future growth in the Internet",* C.S. Seminar, June 1994

[Merit, 94] NIC.MERIT.EDU, *"History of NSFNET growth by networks",* 1 November 1994

[Net Genesis, 95] *Growth of the WWW : servers,* Matthew Gray, Net Genesis, 1995, http://www.netgen.com/info/growth.html

[Network Wizards, 95c] *Growth patterns for WWVV ,* Network Wizards Statistics, November 1995

[Network Wizards, 95a] Network Wizards, *"Internet Domain Survey",* July 1995, http://nw.com

[Network Wizards, 95b] Network Wizards, *"Internet Domain Survey Archives 1980-1995",* ftp.nw.com, July 1995

[Patton, 95] Patton, MA., *"DNS Resource Records for Nimrod routing architecture",* Internet Draft - work in progress, June 1995

[Pesce *et al,* 95] Pesce, M., Kennard, P., Parisi, S., electronic document, *"Cyberspace",* http://vrml.wired.com/concepts/pesce-www.html

[Ramanathan, 95a] Ramanathan, R., *"Multicast support for NIMROD: Requirements and solution approaches",* Internet draft - work in progress, March 1995

[Ramanathan, 95b] Ramanathan, R., *"Mobility support for NIMROD: Requirements and solution approaches",* Internet draft - work in progress, March 1995

[Thomson *et al,* 94] Thomson S., Huitema C., *"DNS Extensions to support Simple Internet Protocol Plus (SIPP)",* working draft, March 1994

[Tsuchiya, 91a] Tsuchiya P. *,"Extending the IP Internet through address reuse",* work in progress, December 1991

[Ullman, 94] Ullman R.L., *"CATNIP: Common architecture for Next-generation Internet Protocol",* working draft, March 1994