               An Architecture for IPv6 Unicast Address Allocation

Status of this Memo

   This document provides information for the Internet community. This
   memo does not specify an Internet standard of any kind. Distribution
   of this memo is unlimited.

Abstract

   This document provides an architecture for allocating IPv6 [1]
   unicast addresses in the Internet. The overall IPv6 addressing
   architecture is defined in [2]. This document does not go into the
   details of an addressing plan.

1.   Scope

   The global internet can be modeled as a collection of hosts
   interconnected via transmission and switching facilities. Control
   over the collection of hosts and the transmission and switching
   facilities that compose the networking resources of the global
   internet is not homogeneous, but is distributed among multiple
   administrative authorities. Resources under control of a single
   administration within a contiguous segment of network topology form a
   domain. For the rest of this paper, 'domain' and 'routing domain'
   will be used interchangeably.

   Domains that share their resources with other domains are called
   network service providers (or just providers). Domains that utilize
   other domain's resources are called network service subscribers (or
   just subscribers). A given domain may act as a provider and a
   subscriber simultaneously.

There are two aspects of interest when discussing IPv6 unicast
address allocation within the Internet. The first is the set of
administrative requirements for obtaining and allocating IPv6
addresses; the second is the technical aspect of such assignments,
having largely to do with routing, both within a routing domain
(intra-domain routing) and between routing domains (inter-domain
routing). This paper focuses on the technical issues.

In the current Internet many routing domains (such as corporate and
campus networks) attach to transit networks (such as regionals) in
only one or a small number of carefully controlled access points.
The former act as subscribers, while the latter act as providers.

Addressing solutions which require substantial changes or constraints
on the current topology are not considered.

The architecture and recommendations in this paper are oriented
primarily toward the large-scale division of IPv6 address allocation
in the Internet. Topics covered include:

    - Benefits of encoding some topological information in IPv6
      addresses to significantly reduce routing protocol overhead;

    - The anticipated need for additional levels of hierarchy in
      Internet addressing to support network growth;

    - The recommended mapping between Internet topological entities
      (i.e., service providers, and service subscribers) and IPv6
      addressing and routing components;

    - The recommended division of IPv6 address assignment among
      service providers (e.g., backbones, regionals), and service
      subscribers (e.g., sites);

    - Allocation of the IPv6 addresses by the Internet Registry;

    - Choice of the high-order portion of the IPv6 addresses in leaf
      routing domains that are connected to more than one service
      provider (e.g., backbone or a regional network).

It is noted that there are other aspects of IPv6 address allocation,
both technical and administrative, that are not covered in this
paper. Topics not covered or mentioned only superficially include:

    - A specific plan for address assignment;

    - Embedding address spaces from other network layer protocols
      (including IPv4) in the IPv6 address space and the addressing

architecture for such embedded addresses;

- Multicast addressing;

- Address allocation for mobile hosts;

- Identification of specific administrative domains in the
  Internet;

- Policy or mechanisms for making registered information known to
  third parties (such as the entity to which a specific IPv6
  address or a potion of the IPv6 address space has been
  allocated);

- How a routing domain (especially a site) should organize its
  internal topology or allocate portions of its IPv6 address
  space; the relationship between topology and addresses is
  discussed, but the method of deciding on a particular topology
  or internal addressing plan is not; and,

- Procedures for assigning host IPv6 addresses.

2.   Background

   Some background information is provided in this section that is
   helpful in understanding the issues involved in IPv6 address
   allocation. A brief discussion of IPv6 routing is provided.

   IPv6 partitions the routing problem into three parts:

   - Routing exchanges between end systems and routers,

   - Routing exchanges between routers in the same routing domain,
     and,

   - Routing among routing domains.

3.   IPv6 Addresses and Routing

   For the purposes of this paper, an IPv6 address prefix is defined as
   an IPv6 address and some indication of the leftmost contiguous
   significant bits within this address portion. Throughout this paper
   IPv6 address prefixes will be represented as X/Y, where X is a prefix
   of an IPv6 address in length greater than or equal to that specified

by Y and Y is the (decimal) number of the leftmost contiguous
significant bits within this address. In the notation, X, the prefix
of an IPv6 address [2] will have trailing insignificant digits
removed. Thus, an IPv6 prefix might appear to be 43DC:0A21:76/40.

When determining an administrative policy for IPv6 address
assignment, it is important to understand the technical consequences.
The objective behind the use of hierarchical routing is to achieve
some level of routing data abstraction, or summarization, to reduce
the cpu, memory, and transmission bandwidth consumed in support of
routing.

While the notion of routing data abstraction may be applied to
various types of routing information, this paper focuses on one
particular type, namely reachability information. Reachability
information describes the set of reachable destinations. Abstraction
of reachability information dictates that IPv6 addresses be assigned
according to topological routing structures. However in practice
administrative assignment falls along organizational or political
boundaries. These may not be congruent to topological boundaries and
therefore the requirements of the two may collide. It is necessary to
find a balance between these two needs.

Reachability information abstraction occurs at the boundary between
hierarchically arranged topological routing structures. An element
lower in the hierarchy reports summary reachability information to
its parent(s).

At routing domain boundaries, IPv6 address information is exchanged
(statically or dynamically) with other routing domains. If IPv6
addresses within a routing domain are all drawn from non-contiguous
IPv6 address spaces (allowing no abstraction), then the address
information exchanged at the boundary consists of an enumerated list
of all the IPv6 addresses.

Alternatively, should the routing domain draw IPv6 addresses for all
the hosts within the domain from a single IPv6 address prefix,
boundary routing information can be summarized into the single IPv6
address prefix. This permits substantial data reduction and allows
better scaling (as compared to the uncoordinated addressing discussed
in the previous paragraph).

If routing domains are interconnected in a more-or-less random (i.e.,
non-hierarchical) scheme, it is quite likely that no further
abstraction of routing data can occur. Since routing domains would
have no defined hierarchical relationship, administrators would not
be able to assign IPv6 addresses within the domains out of some
common prefix for the purpose of data abstraction. The result would

be flat inter-domain routing; all routing domains would need explicit
knowledge of all other routing domains that they route to. This can
work well in small and medium sized internets. However, this does
not scale to very large internets. For example, we expect IPv6 to
grow to hundreds of thousands of routing domains in North America
alone. This requires a greater degree of the reachability
information abstraction beyond that which can be achieved at the
'routing domain' level.

In the Internet, it should be possible to significantly constrain the
volume and the complexity of routing information by taking advantage
of the existing hierarchical interconnectivity. This is discussed
further in Section 5. Thus, there is the opportunity for a group of
routing domains each to be assigned an address prefix from a shorter
prefix assigned to another routing domain whose function is to
interconnect the group of routing domains. Each member of the group
of routing domains now has its (somewhat longer) prefix, from which
it assigns its addresses.

The most straightforward case of this occurs when there is a set of
routing domains which are all attached to a single service provider
domain (e.g., regional network), and which use that provider for all
external (inter-domain) traffic. A short prefix may be given to the
provider, which then gives slightly longer prefixes (based on the
provider's prefix) to each of the routing domains that it
interconnects. This allows the provider, when informing other routing
domains of the addresses that it can reach, to abstract the
reachability information for a large number of routing domains into a
single prefix. This approach therefore can allow a great deal of
reduction of routing information, and thereby can greatly improve the
scalability of inter-domain routing.

Clearly, this approach is recursive and can be carried through
several iterations. Routing domains at any 'level' in the hierarchy
may use their prefix as the basis for subsequent suballocations,
assuming that the IPv6 addresses remain within the overall length and
structure constraints.

At this point, we observe that the number of nodes at each lower
level of a hierarchy tends to grow exponentially. Thus the greatest
gains in the reachability information abstraction (for the benefit of
all higher levels of the hierarchy) occur when the reachability
information aggregation occurs near the leaves of the hierarchy; the
gains drop significantly at each higher level. Therefore, the law of
diminishing returns suggests that at some point data abstraction
ceases to produce significant benefits. Determination of the point
at which data abstraction ceases to be of benefit requires a careful
consideration of the number of routing domains that are expected to

occur at each level of the hierarchy (over a given period of time),
compared to the number of routing domains and address prefixes that
can conveniently and efficiently be handled via dynamic inter-domain
routing protocols.

3.1 Efficiency versus Decentralized Control.

If the Internet plans to support a decentralized address
administration, then there is a balance that must be sought between
the requirements on IPv6 addresses for efficient routing and the need
for decentralized address administration. A coherent addressing plan
at any level within the Internet must take the alternatives into
careful consideration.

As an example of administrative decentralization, suppose the IPv6
address prefix 43/8 identifies part of the IPv6 address space
allocated for North America. All addresses within this prefix may be
allocated along topological boundaries in support of increased data
abstraction. Within this prefix, addresses may be allocated on a
per-provider bases, based on geography or some other topologically
significant criteria. For the purposes of this example, suppose that
this prefix is allocated on a per-provider basis. Subscribers within
North America use parts of the IPv6 address space that is underneath
the IPv6 address space of their service providers. Within a routing
domain addresses for subnetworks and hosts are allocated from the
unique IPv6 prefix assigned to the domain according to the addressing
plan for that domain.

4.   IPv6 Address Administration and Routing in the Internet

Internet routing components -- service providers (e.g., backbones,
regional networks), and service subscribers (e.g., sites or campuses)
-- are arranged hierarchically for the most part. A natural mapping
from these components to IPv6 routing components is for providers and
subscribers to act as routing domains.

Alternatively, a subscriber (e.g., a site) may choose to operate as a
part of a domain formed by a service provider. We assume that some,
if not most, sites will prefer to operate as part of their provider's
routing domain, exchanging routing information directly with the
provider. The site is still allocated a prefix from the provider's
address space, and the provider will advertise its own prefix into
inter-domain routing.

Given such a mapping, where should address administration and
allocation be performed to satisfy both administrative
decentralization and data abstraction? The following possibilities
are considered:

   1) At some part within a routing domain,

   2) At the leaf routing domain,

   3) At the transit routing domain (TRD), and

   4) At some other, more general boundaries, such as at the
      continental boundary.

A part within a routing domain corresponds to any arbitrary connected
set of subnetworks. If a domain is composed of multiple subnetworks,
they are interconnected via routers. Leaf routing domains correspond
to sites, where the primary purpose is to provide intra-domain
routing services. Transit routing domains are deployed to carry
transit (i.e., inter-domain) traffic; backbones and providers are
TRDs. More general boundaries can be seen as topologically
significant collections of TRDs.

The greatest burden in transmitting and operating on reachability
information is at the top of the routing hierarchy, where
reachability information tends to accumulate. In the Internet, for
example, providers must manage reachability information for all
subscribers directly connected to the provider. Traffic destined for
other providers is generally routed to the backbones (which act as
providers as well). The backbones, however, must be cognizant of the
reachability information for all attached providers and their
associated subscribers.

In general, the advantage of abstracting routing information at a
given level of the routing hierarchy is greater at the higher levels
of the hierarchy. There is relatively little direct benefit to the
administration that performs the abstraction, since it must maintain
routing information individually on each attached topological routing
structure.

For example, suppose that a given site is trying to decide whether to
obtain an IPv6 address prefix directly from the IPv6 address space
allocated for North America, or from the IPv6 address space allocated
to its service provider. If considering only their own self-interest,
the site itself and the attached provider have little reason to
choose one approach or the other. The site must use one prefix or
another; the source of the prefix has little effect on routing
efficiency within the site. The provider must maintain information

about each attached site in order to route, regardless of any
commonality in the prefixes of the sites.

However, there is a difference when the provider distributes routing
information to other providers (e.g., backbones or TRDs). In the
first case, the provider cannot aggregate the site's address into its
own prefix; the address must be explicitly listed in routing
exchanges, resulting in an additional burden to other providers which
must exchange and maintain this information.

In the second case, each other provider (e.g., backbone or TRD) sees
a single address prefix for the provider, which encompasses the new
site. This avoids the exchange of additional routing information to
identify the new site's address prefix. Thus, the advantages
primarily accrue to other providers which maintain routing
information about this site and provider.

One might apply a supplier/consumer model to this problem: the higher
level (e.g., a backbone) is a supplier of routing services, while the
lower level (e.g., a TRD) is the consumer of these services. The
price charged for services is based upon the cost of providing them.
The overhead of managing a large table of addresses for routing to an
attached topological entity contributes to this cost.

At present the Internet, however, is not a market economy. Rather,
efficient operation is based on cooperation. The recommendations
discussed below describe simple and tractable ways of managing the
IPv6 address space that benefit the entire community.

4.1   Administration of IPv6 addresses within a domain.

If individual hosts take their IPv6 addresses from a myriad of
unrelated IPv6 address spaces, there will be effectively no data
abstraction beyond what is built into existing intra-domain routing
protocols. For example, assume that within a routing domain uses
three independent prefixes assigned from three different IPv6 address
spaces associated with three different attached providers.

This has a negative effect on inter-domain routing, particularly on
those other domains which need to maintain routes to this domain.
There is no common prefix that can be used to represent these IPv6
addresses and therefore no summarization can take place at the
routing domain boundary. When addresses are advertised by this
routing domain to other routing domains, an enumerated list of the
three individual prefixes must be used.

The number of IPv6 prefixes that leaf routing domains would advertise
is on the order of the number of prefixes assigned to the domain; the
number of prefixes a provider's routing domain would advertise is
approximately the number of prefixes attached to the client leaf
routing domains; and for a backbone this would be summed across all
attached providers. This situation is just barely acceptable in the
current Internet, and is intractable for the IPv6 Internet. A
greater degree of hierarchical information reduction is necessary to
allow continued growth in the Internet.

4.2   Administration at the Leaf Routing Domain

As mentioned previously, the greatest degree of data abstraction
comes at the lowest levels of the hierarchy. Providing each leaf
routing domain (that is, site) with a contiguous block of addresses
from its provider's address block results in the biggest single
increase in abstraction. From outside the leaf routing domain, the
set of all addresses reachable in the domain can then be represented
by a single prefix. Further, all destinations reachable within the
provider's prefix can be represented by a single prefix.

For example, consider a single campus which is a leaf routing domain
which would currently require 4 different IPv6 prefixes. Instead,
they may be given a single prefix which provides the same number of
destination addresses. Further, since the prefix is a subset of the
provider's prefix, they impose no additional burden on the higher
levels of the routing hierarchy.

There is a close relationship between hosts and routing domains. The
routing domain represents the only path between a host and the rest
of the internetwork. It is reasonable that this relationship also
extend to include a common IPv6 addressing space. Thus, the hosts
within the leaf routing domain should take their IPv6 addresses from
the prefix assigned to the leaf routing domain.

4.3   Administration at the Transit Routing Domain

Two kinds of transit routing domains are considered, direct providers
and indirect providers. Most of the subscribers of a direct provider
are domains that act solely as service subscribers (they carry no
transit traffic). Most of the subscribers of an indirect provider are
domains that, themselves, act as service providers. In present
terminology a backbone is an indirect provider, while an NSFnet
regional is an example of a direct provider. Each case is discussed

separately below.

4.3.1   Direct Service Providers

In a provider-based addressing plan, direct service providers should
use their IPv6 address space for assigning IPv6 addresses from a
unique prefix to the leaf routing domains that they serve. The
benefits derived from data abstraction are greater than in the case
of leaf routing domains, and the additional degree of data
abstraction provided by this may be necessary in the short term.

As an illustration consider an example of a direct provider that
serves 100 clients. If each client takes its addresses from 4
independent address spaces then the total number of entries that are
needed to handle routing to these clients is 400 (100 clients times 4
providers). If each client takes its addresses from a single address
space then the total number of entries would be only 100. Finally, if
all the clients take their addresses from the same address space then
the total number of entries would be only 1.

We expect that in the near term the number of routing domains in the
Internet will grow to the point that it will be infeasible to route
on the basis of a flat field of routing domains. It will therefore be
essential to provide a greater degree of information abstraction with
IPv6.

Direct providers may give part of their address space (prefixes) to
leaf domains, based on an address prefix given to the provider. This
results in direct providers advertising to other providers a small
fraction of the number of address prefixes that would be necessary if
they enumerated the individual prefixes of the leaf routing domains.
This represents a significant savings given the expected scale of
global internetworking.

The efficiencies gained in inter-domain routing clearly warrant the
adoption of IPv6 address prefixes derived from the IPv6 address space
of the providers.

The mechanics of this scenario are straightforward. Each direct
provider is given a unique small set of IPv6 address prefixes, from
which its attached leaf routing domains can allocate slightly longer
IPv6 address prefixes. For example assume that NIST is a leaf
routing domain whose inter-domain link is via SURANet. If SURANet is
assigned an unique IPv6 address prefix 43DC:0A21/32, NIST could use a
unique IPv6 prefix of 43DC:0A21:7652:34/56.

If a direct service provider is connected to another provider(s)
(either direct or indirect) via multiple attachment points, then in
certain cases it may be advantageous to the direct provider to exert
a certain degree of control over the coupling between the attachment
points and flow of the traffic destined to a particular subscriber.
Such control can be facilitated by first partitioning all the
subscribers into groups, such that traffic destined to all the
subscribers within a group should flow through a particular
attachment point. Once the partitioning is done, the address space of
the provider is subdivided along the group boundaries. A leaf routing
domain that is willing to accept prefixes derived from its direct
provider gets a prefix from the provider's address space subdivision
associated with the group the domain belongs to.

At the attachment point (between the direct and indirect providers)
the direct provider advertises both an address prefix that
corresponds to the address space of the provider, and one or more
address prefixes that correspond to the address space associated with
each subdivision. The latter prefixes match the former prefix, but
are longer than the former prefix. Use of the "longest match"
forwarding algorithm by the recipients of these prefixes (e.g., a
router within the indirect provider) results in forcing the flow of
the traffic to destinations depicted by the longer address prefixes
through the attachment point where these prefixes are advertised to
the indirect provider.

For example, assume that SURANet is connected to another regional
provider, NEARNet, at two attachment points, A1 and A2. SURANet is
assigned a unique IPv6 address prefix 43DC:0A21/32. To exert control
over the traffic flow destined to a particular subscriber within
SURANet, SURANet may subdivide the address space assigned to it into
two groups, 43DC:0A21:8/34 and 43DC:0A21:C/34. The former group may
be used for sites attached to SURANet that are closer (as determined
by the topology within SURANet) to A1, while the latter group may be
used for sites that are closer to A2. The SURANet router at A1
advertises both 43DC:0A21/32 and 43DC:0A21:8/34 address prefixes to
the router in NEARNet. Likewise, the SURANet router at A2 advertises
both 43DC:0A21/32 and 43DC:0A21:C/34 address prefixes to the router
in NEARNet. Traffic that flows through NEARNet to destinations that
match 43DC:0A21:8/34 address prefix would enter SURANet at A1, while
traffic to destinations that match 43DC:0A21:C/34 address prefix
would enter SURANet at A2.

Note that the advertisement by the direct provider of the routing
information associated with each subdivision must be done with care
to ensure that such an advertisement would not result in a global
distribution of separate reachability information associated with
each subdivision, unless such distribution is warranted for some

other purposes (e.g., supporting certain aspects of policy-based
routing).

4.3.2   Indirect Providers (Backbones)

There does not at present appear to be a strong case for direct
providers to take their address spaces from the the IPv6 space of an
indirect provider (e.g., backbone). The benefit in routing data
abstraction is relatively small. The number of direct providers today
is in the tens and an order of magnitude increase would not cause an
undue burden on the backbones. Also, it may be expected that as time
goes by there will be increased direct interconnection of the direct
providers, leaf routing domains directly attached to the backbones,
and international links directly attached to the providers. Under
these circumstances, the distinction between direct and indirect
providers may become blurred.

An additional factor that discourages allocation of IPv6 addresses
from a backbone prefix is that the backbones and their attached
providers are perceived as being independent. Providers may take
their long-haul service from one or more backbones, or may switch
backbones should a more cost-effective service be provided elsewhere.
Having IPv6 addresses derived from a backbone is inconsistent with
the nature of the relationship.

4.4   Multi-homed Routing Domains

The discussions in Section 4.3 suggest methods for allocating IPv6
addresses based on direct or indirect provider connectivity. This
allows a great deal of information reduction to be achieved for those
routing domains which are attached to a single TRD. In particular,
such routing domains may select their IPv6 addresses from a space
delegated to them by the direct provider. This allows the provider,
when announcing the addresses that it can reach to other providers,
to use a single address prefix to describe a large number of IPv6
addresses corresponding to multiple routing domains.

However, there are additional considerations for routing domains
which are attached to multiple providers. Such 'multi-homed' routing
domains may, for example, consist of single-site campuses and
companies which are attached to multiple backbones, large
organizations which are attached to different providers at different
locations in the same country, or multi-national organizations which
are attached to backbones in a variety of countries worldwide. There

are a number of possible ways to deal with these multi-homed routing
domains.

4.4.1 Solution 1

One possible solution is for each multi-homed organization to obtain
its IPv6 address space independently of the providers to which it is
attached. This allows each multi-homed organization to base its IPv6
assignments on a single prefix, and to thereby summarize the set of
all IPv6 addresses reachable within that organization via a single
prefix. The disadvantage of this approach is that since the IPv6
address for that organization has no relationship to the addresses of
any particular TRD, the TRDs to which this organization is attached
will need to advertise the prefix for this organization to other
providers. Other providers (potentially worldwide) will need to
maintain an explicit entry for that organization in their routing
tables.

For example, suppose that a very large North American company 'Mega
Big International Incorporated' (MBII) has a fully interconnected
internal network and is assigned a single prefix as part of the North
American prefix. It is likely that outside of North America, a
single entry may be maintained in routing tables for all North
American Destinations. However, within North America, every provider
will need to maintain a separate address entry for MBII. If MBII is
in fact an international corporation, then it may be necessary for
every provider worldwide to maintain a separate entry for MBII
(including backbones to which MBII is not attached). Clearly this may
be acceptable if there are a small number of such multi-homed routing
domains, but would place an unacceptable load on routers within
backbones if all organizations were to choose such address
assignments. This solution may not scale to internets where there
are many hundreds of thousands of multi-homed organizations.

4.4.2 Solution 2

A second possible approach would be for multi-homed organizations to
be assigned a separate IPv6 address space for each connection to a
TRD, and to assign a single prefix to some subset of its domain(s)
based on the closest interconnection point. For example, if MBII had
connections to two providers in the U.S. (one east coast, and one
west coast), as well as three connections to national backbones in
Europe, and one in the far east, then MBII may make use of six
different address prefixes. Each part of MBII would be assigned a

single address prefix based on the nearest connection.

For purposes of external routing of traffic from outside MBII to a
destination inside of MBII, this approach works similarly to treating
MBII as six separate organizations. For purposes of internal routing,
or for routing traffic from inside of MBII to a destination outside
of MBII, this approach works the same as the first solution.

If we assume that incoming traffic (coming from outside of MBII, with
a destination within MBII) is always to enter via the nearest point
to the destination, then each TRD which has a connection to MBII
needs to announce to other TRDs the ability to reach only those parts
of MBII whose address is taken from its own address space. This
implies that no additional routing information needs to be exchanged
between TRDs, resulting in a smaller load on the inter-domain routing
tables maintained by TRDs when compared to the first solution. This
solution therefore scales better to extremely large internets
containing very large numbers of multi-homed organizations.

One problem with the second solution is that backup routes to multi-
homed organizations are not automatically maintained. With the first
solution, each TRD, in announcing the ability to reach MBII,
specifies that it is able to reach all of the hosts within MBII. With
the second solution, each TRD announces that it can reach all of the
hosts based on its own address prefix, which only includes some of
the hosts within MBII. If the connection between MBII and one
particular TRD were severed, then the hosts within MBII with
addresses based on that TRD would become unreachable via inter-domain
routing. The impact of this problem can be reduced somewhat by
maintenance of additional information within routing tables, but this
reduces the scaling advantage of the second approach.

The second solution also requires that when external connectivity
changes, internal addresses also change.

Also note that this and the previous approach will tend to cause
packets to take different routes. With the first approach, packets
from outside of MBII destined for within MBII will tend to enter via
the point which is closest to the source (which will therefore tend
to maximize the load on the networks internal to MBII). With the
second solution, packets from outside destined for within MBII will
tend to enter via the point which is closest to the destination
(which will tend to minimize the load on the networks within MBII,
and maximize the load on the TRDs).

These solutions also have different effects on policies. For example,
suppose that country 'X' has a law that traffic from a source within
country X to a destination within country X must at all times stay

entirely within the country. With the first solution, it is not
possible to determine from the destination address whether or not the
destination is within the country. With the second solution, a
separate address may be assigned to those hosts which are within
country X, thereby allowing routing policies to be followed.
Similarly, suppose that 'Little Small Company' (LSC) has a policy
that its packets may never be sent to a destination that is within
MBII. With either solution, the routers within LSC may be configured
to discard any traffic that has a destination within MBII's address
space. However, with the first solution this requires one entry; with
the second it requires many entries and may be impossible as a
practical matter.

4.4.3 Solution 3

There are other possible solutions as well. A third approach is to
assign each multi-homed organization a single address prefix, based
on one of its connections to a TRD. Other TRDs to which the multi-
homed organization are attached maintain a routing table entry for
the organization, but are extremely selective in terms of which other
TRDs are told of this route. This approach will produce a single
'default' routing entry which all TRDs will know how to reach (since
presumably all TRDs will maintain routes to each other), while
providing more direct routing in some cases.

There is at least one situation in which this third approach is
particularly appropriate. Suppose that a special interest group of
organizations have deployed their own provider. For example, lets
suppose that the U.S. National Widget Manufacturers and Researchers
have set up a U.S.-wide provider, which is used by corporations who
manufacture widgets, and certain universities which are known for
their widget research efforts. We can expect that the various
organizations which are in the widget group will run their internal
networks as separate routing domains, and most of them will also be
attached to other TRDs (since most of the organizations involved in
widget manufacture and research will also be involved in other
activities). We can therefore expect that many or most of the
organizations in the widget group are dual-homed, with one attachment
for widget-associated communications and the other attachment for
other types of communications. Let's also assume that the total
number of organizations involved in the widget group is small enough
that it is reasonable to maintain a routing table containing one
entry per organization, but that they are distributed throughout a
larger internet with many millions of (mostly not widget-associated)
routing domains.

With the third approach, each multi-homed organization in the widget
group would make use of an address assignment based on its other
attachment(s) to TRDs (the attachments not associated with the widget
group). The widget provider would need to maintain routes to the
routing domains associated with the various member organizations.
Similarly, all members of the widget group would need to maintain a
table of routes to the other members via the widget provider.
However, since the widget provider does not inform other general
worldwide TRDs of what addresses it can reach (since the provider is
not intended for use by other outside organizations), the relatively
large set of routing prefixes needs to be maintained only in a
limited number of places. The addresses assigned to the various
organizations which are members of the widget group would provide a
'default route' via each members other attachments to TRDs, while
allowing communications within the widget group to use the preferred
path.

4.4.4 Solution 4

A fourth solution involves assignment of a particular address prefix
for routing domains which are attached to precisely two (or more)
specific routing domains. For example, suppose that there are two
providers 'SouthNorthNet' and 'NorthSouthNet' which have a very large
number of customers in common (i.e., there are a large number of
routing domains which are attached to both). Rather than getting two
address prefixes these organizations could obtain three prefixes.
Those routing domains which are attached to NorthSouthNet but not
attached to SouthNorthNet obtain an address assignment based on one
of the prefixes. Those routing domains which are attached to
SouthNorthNet but not to NorthSouthNet would obtain an address based
on the second prefix. Finally, those routing domains which are
multi-homed to both of these networks would obtain an address based
on the third prefix. Each of these two TRDs would then advertise two
prefixes to other TRDs, one prefix for leaf routing domains attached
to it only, and one prefix for leaf routing domains attached to both.

This fourth solution is likely to be important when use of public
data networks becomes more common. In particular, it is likely that
at some point in the future a substantial percentage of all routing
domains will be attached to public data networks. In this case,
nearly all government-sponsored networks (such as some current
regionals) may have a set of customers which overlaps substantially
with the public networks.

4.4.5 Summary

   There are therefore a number of possible solutions to the problem of
   assigning IPv6 addresses to multi-homed routing domains. Each of
   these solutions has very different advantages and disadvantages.
   Each solution places a different real (i.e., financial) cost on the
   multi-homed organizations, and on the TRDs (including those to which
   the multi-homed organizations are not attached).

   In addition, most of the solutions described also highlight the need
   for each TRD to develop a policy on whether and under what conditions
   to accept addresses that are not based on its own address prefix, and
   how such non-local addresses will be treated. For example, a somewhat
   conservative policy might be that non-local IPv6 address prefixes
   will be accepted from any attached leaf routing domain, but not
   advertised to other TRDs. In a less conservative policy, a TRD might
   accept such non-local prefixes and agree to exchange them with a
   defined set of other TRDs (this set could be an a priori group of
   TRDs that have something in common such as geographical location, or
   the result of an agreement specific to the requesting leaf routing
   domain). Various policies involve real costs to TRDs, which may be
   reflected in those policies.

4.5   Private Links

   The discussion up to this point concentrates on the relationship
   between IPv6 addresses and routing between various routing domains
   over transit routing domains, where each transit routing domain
   interconnects a large number of routing domains and offers a more-
   or-less public service.

   However, there may also exist a number of links which interconnect
   two routing domains in such a way, that usage of these links may be
   limited to carrying traffic only between the two routing domains.
   We'll refer to such links as "private".

   For example, let's suppose that the XYZ corporation does a lot of
   business with MBII. In this case, XYZ and MBII may contract with a
   carrier to provide a private link between the two corporations, where
   this link may only be used for packets whose source is within one of
   the two corporations, and whose destination is within the other of
   the two corporations. Finally, suppose that the point-to-point link
   is connected between a single router (router X) within XYZ
   corporation and a single router (router M) within MBII. It is
   therefore necessary to configure router X to know which addresses can

be reached over this link (specifically, all addresses reachable in
MBII). Similarly, it is necessary to configure router M to know which
addresses can be reached over this link (specifically, all addresses
reachable in XYZ Corporation).

The important observation to be made here is that the additional
connectivity due to such private links may be ignored for the purpose
of IPv6 address allocation, and do not pose a problem for routing on
a larger scale. This is because the routing information associated
with such connectivity is not propagated throughout the internet, and
therefore does not need to be collapsed into a TRD's prefix.

In our example, let's suppose that the XYZ corporation has a single
connection to a regional, and has therefore uses the IPv6 address
space from the space given to that regional. Similarly, let's
suppose that MBII, as an international corporation with connections
to six different providers, has chosen the second solution from
Section 4.4, and therefore has obtained six different address
allocations. In this case, all addresses reachable in the XYZ
Corporation can be described by a single address prefix (implying
that router M only needs to be configured with a single address
prefix to represent the addresses reachable over this link). All
addresses reachable in MBII can be described by six address prefixes
(implying that router X needs to be configured with six address
prefixes to represent the addresses reachable over the link).

In some cases, such private links may be permitted to forward traffic
for a small number of other routing domains, such as closely
affiliated organizations. This will increase the configuration
requirements slightly. However, provided that the number of
organizations using the link is relatively small, then this still
does not represent a significant problem.

Note that the relationship between routing and IPv6 addressing
described in other sections of this paper is concerned with problems
in scaling caused by large, essentially public transit routing
domains which interconnect a large number of routing domains.
However, for the purpose of IPv6 address allocation, private links
which interconnect only a small number of private routing domains do
not pose a problem, and may be ignored. For example, this implies
that a single leaf routing domain which has a single connection to a
'public' provider (e.g., the Alternet), plus a number of private
links to other leaf routing domains, can be treated as if it were
single-homed to the provider for the purpose of IPv6 address
allocation. We expect that this is also another way of dealing with
multi-homed domains.

4.6   Zero-Homed Routing Domains

   Currently, a very large number of organizations have internal
   communications networks which are not connected to any service
   providers. Such organizations may, however, have a number of private
   links that they use for communications with other organizations. Such
   organizations do not participate in global routing, but are satisfied
   with reachability to those organizations with which they have
   established private links. These are referred to as zero-homed
   routing domains.

   Zero-homed routing domains can be considered as the degenerate case
   of routing domains with private links, as discussed in the previous
   section, and do not pose a problem for inter-domain routing. As
   above, the routing information exchanged across the private links
   sees very limited distribution, usually only to the routing domain at
   the other end of the link. Thus, there are no address abstraction
   requirements beyond those inherent in the address prefixes exchanged
   across the private link.

   However, it is important that zero-homed routing domains use valid
   globally unique IPv6 addresses. Suppose that the zero-homed routing
   domain is connected through a private link to a routing domain.
   Further, this routing domain participates in an internet that
   subscribes to the global IPv6 addressing plan. This domain must be
   able to distinguish between the zero-homed routing domain's IPv6
   addresses and any other IPv6 addresses that it may need to route to.
   The only way this can be guaranteed is if the zero-homed routing
   domain uses globally unique IPv6 addresses.

   Whereas globally unique addresses are necessary to differentiate
   between destinations in the Internet, globally unique addresses may
   not be sufficient to guarantee global connectivity. If a zero-homed
   routing domain gets connected to the Internet, the block of addresses
   used within the domain may not be related to the block of addresses
   allocated to the domain's direct provider. In order to maintain the
   gains given by hierarchical routing and address assignment the zero-
   homed domain should under such circumstances change addresses
   assigned to the systems within the domain.

4.7   Continental aggregation

   Another level of hierarchy may also be used in this addressing scheme
   to further reduce the amount of routing information necessary for

global routing. Continental aggregation is useful because
continental boundaries provide natural barriers to topological
connection and administrative boundaries. Thus, it presents a
natural boundary for another level of aggregation of inter-domain
routing information. To make use of this, it is necessary that each
continent be assigned an appropriate contiguous block of addresses.
Providers (both direct and indirect) within that continent would
allocate their addresses from this space. Note that there are
numerous exceptions to this, in which a service provider (either
direct or indirect) spans a continental division. These exceptions
can be handled similarly to multi-homed routing domains, as discussed
above.

The benefit of continental aggregation is that it helps to absorb the
entropy introduced within continental routing caused by the cases
where an organization must use an address prefix which must be
advertised beyond its direct provider. In such cases, if the address
is taken from the continental prefix, the additional cost of the
route is not propagated past the point where continental aggregation
takes place.

Note that, in contrast to the case of providers, the aggregation of
continental routing information may not be done on the continent to
which the prefix is allocated. The cost of inter-continental links
(and especially trans-oceanic links) is very high. If aggregation is
performed on the 'near' side of the link, then routing information
about unreachable destinations within that continent can only reside
on that continent. Alternatively, if continental aggregation is done
on the 'far' side of an inter-continental link, the 'far' end can
perform the aggregation and inject it into continental routing. This
means that destinations which are part of the continental
aggregation, but for which there is not a corresponding more specific
prefix can be rejected before leaving the continent on which they
originated.

For example, suppose that Europe is assigned a prefix of 46/8, such
that European routing also contains the longer prefixes 46DC:0A01/32
and 46DC:0A02/32 . All of the longer European prefixes may be
advertised across a trans-Atlantic link to North America. The router
in North America would then aggregate these routes, and only
advertise the prefix 46/8 into North American routing. Packets which
are destined for 46DC:0A01:1234:5678:ABCD:8765:4321:AABB would
traverse North American routing, but would encounter the North
American router which performed the European aggregation. If the
prefix 46DC:0A01/32 is unreachable, the router would drop the packet
and send an unreachable message without using the trans-Atlantic
link.

4.8   Private (Local Use) Addresses

   Many domains will have hosts which, for one reason or another, will
   not require globally unique IPv6 addresses. A domain which decides
   to use IPv6 addresses out of the private address space is able to do
   so without address allocation from any authority. Conversely, the
   private address prefix need not be advertised throughout the public
   portion of the Internet.

   In order to use private address space, a domain needs to determine
   which hosts do not need to have network layer connectivity outside
   the domain in the foreseeable future. Such hosts will be called
   private hosts, and may use the private addresses described above if
   it is topologically convenient. Private hosts can communicate with
   all other hosts inside the domain, both public and private. However,
   they cannot have IPv6 connectivity to any external host. While not
   having external network layer connectivity, a private host can still
   have access to external services via application layer relays.
   Public hosts do not have connectivity to private hosts outside of
   their own domain.

   Because private addresses have no global meaning, reachability
   information associated with the private address space shall not be
   propagated on inter-domain links, and packets with private source or
   destination addresses should not be forwarded across such links.
   Routers in domains not using private address space, especially those
   of Internet service providers, are expected to be configured to
   reject (filter out) routing information that carries reachability
   information associated with private addresses. If such a router
   receives such information the rejection shall not be treated as a
   routing protocol error.

   In addition, indirect references to private addresses should be
   contained within the enterprise that uses these addresses. Prominent
   example of such references are DNS Resource Records. A possible
   approach to avoid leaking of DNS RRs is to run two nameservers, one
   external server authoritative for all globally unique IP addresses of
   the enterprise and one internal nameserver authoritative for all IP
   addresses of the enterprise, both public and private. In order to
   ensure consistency both these servers should be configured from the
   same data of which the external nameserver only receives a filtered
   version. The resolvers on all internal hosts, both public and
   private, query only the internal nameserver. The external server
   resolves queries from resolvers outside the enterprise and is linked
   into the global DNS. The internal server forwards all queries for
   information outside the enterprise to the external nameserver, so all
   internal hosts can access the global DNS. This ensures that

information about private hosts does not reach resolvers and
nameservers outside the enterprise.

4.9   Interaction with Policy Routing

We assume that any inter-domain routing protocol will have difficulty
trying to aggregate multiple destinations with dissimilar policies.
At the same time, the ability to aggregate routing information while
not violating routing policies is essential. Therefore, we suggest
that address allocation authorities attempt to allocate addresses so
that aggregates of destinations with similar policies can be easily
formed.

5.   Recommendations

We anticipate that the current exponential growth of the Internet
will continue or accelerate for the foreseeable future. In addition,
we anticipate a rapid internationalization of the Internet. The
ability of routing to scale is dependent upon the use of data
abstraction based on hierarchical IPv6 addresses. It is therefore
essential to choose a hierarchical structure for IPv6 addresses with
great care.

Great attention must be paid to the definition of addressing
structures to balance the conflicting goals of:

  - Route optimality

  - Routing algorithm efficiency

  - Ease and administrative efficiency of address registration

  - Considerations for what addresses are assigned by what addressing
     authority

It is in the best interests of the internetworking community that the
cost of operations be kept to a minimum where possible. In the case
of IPv6 address allocation, this again means that routing data
abstraction must be encouraged.

In order for data abstraction to be possible, the assignment of IPv6
addresses must be accomplished in a manner which is consistent with
the actual physical topology of the Internet. For example, in those
cases where organizational and administrative boundaries are not

related to actual network topology, address assignment based on such
organization boundaries is not recommended.

The intra-domain routing protocols allow for information abstraction
to be maintained within a domain. For zero-homed and single-homed
routing domains (which are expected to remain zero-homed or single-
homed), we recommend that the IPv6 addresses assigned within a single
routing domain use a single address prefix assigned to that domain.
Specifically, this allows the set of all IPv6 addresses reachable
within a single domain to be fully described via a single prefix.

We anticipate that the total number of routing domains existing on a
worldwide Internet to be great enough that additional levels of
hierarchical data abstraction beyond the routing domain level will be
necessary.

In most cases, network topology will have a close relationship with
national boundaries. For example, the degree of network connectivity
will often be greater within a single country than between countries.
It is therefore appropriate to make specific recommendations based on
national boundaries, with the understanding that there may be
specific situations where these general recommendations need to be
modified.

Further, from experience with IPv4, we feel that continental
aggregation is beneficial and should be supported where possible as a
means of limiting the unwarranted spread of routing exceptions.

5.1   Recommendations for an address allocation plan

We anticipate that public interconnectivity between private routing
domains will be provided by a diverse set of TRDs, including (but not
necessarily limited to):

   - Backbone networks;

   - A number of regional or national networks; and,

   - A number of commercial Public Data Networks.

These networks will not be interconnected in a strictly hierarchical
manner (for example, there is expected to be direct connectivity
between regionals, and all of these types of networks may have direct
international connections). These TRDs will be used to interconnect
a wide variety of routing domains, each of which may comprise a
single corporation, part of a corporation, a university campus, a

government agency, or other organizational unit.

In addition, some private corporations may be expected to make use of
dedicated private TRDs for communication within their own
corporation.

We anticipate that the great majority of routing domains will be
attached to only one of the TRDs. This will permit hierarchical
address aggregation based on TRD. We therefore strongly recommend
that addresses be assigned hierarchically, based on address prefixes
assigned to individual TRDs.

To support continental aggregation of routes, we recommend that all
addresses for TRDs which are wholly within a continent be taken from
the continental prefix.

For the proposed address allocation scheme, this implies that
portions of IPv6 address space should be assigned to each TRD
(explicitly including the backbones and regionals). For those leaf
routing domains which are connected to a single TRD, they should be
assigned a prefix value from the address space assigned to that TRD.

For routing domains which are not attached to any publically
available TRD, there is not the same urgent need for hierarchical
address aggregation. We do not, therefore, make any additional
recommendations for such 'isolated' routing domains. Where such
domains are connected to other domains by private point-to-point
links, and where such links are used solely for routing between the
two domains that they interconnect, again no additional technical
problems relating to address abbreviation is caused by such a link,
and no specific additional recommendations are necessary. We do
recommend that since such domains may wish to use a private address
space, that the addressing plan specify a specific prefix for private
addressing.

Further, in order to allow aggregation of IPv6 addresses at national
and continental boundaries into as few prefixes as possible, we
further recommend that IPv6 addresses allocated to routing domains
should be assigned based on each routing domain's connectivity to
national and continental Internet backbones.

5.2   Recommendations for Multi-Homed Routing Domains

Some routing domains will be attached to multiple TRDs within the
same country, or to TRDs within multiple different countries. We
refer to these as 'multi-homed' routing domains. Clearly the strict

hierarchical model discussed above does not neatly handle such
routing domains.

There are several possible ways that these multi-homed routing
domains may be handled, as described in Section 4.4. Each of these
methods vary with respect to the amount of information that must be
maintained for inter-domain routing and also with respect to the
inter-domain routes. In addition, the organization that will bear the
brunt of this cost varies with the possible solutions. For example,
the solutions vary with respect to:

   - Resources used within routers within the TRDs;

   - Administrative cost on TRD personnel; and,

   - Difficulty of configuration of policy-based inter-domain routing
      information within leaf routing domains.

Also, the solution used may affect the actual routes which packets
follow, and may effect the availability of backup routes when the
primary route fails.

For these reasons it is not possible to mandate a single solution for
all situations. Rather, economic considerations will require a
variety of solutions for different routing domains, service
providers, and backbones.

6.   Security Considerations

Security issues are not discussed in this document.

7.   Acknowledgments

This document is largely based on RFC 1518. The section on Private
Addresses borrowed heavily from RFC 1597.

We'd like to thank Havard Eidnes, Bill Manning, Roger Fajman for
their review and constructive comments.

REFERENCES

   [1] Deering, S., and R. Hinden, Editors, "Internet Protocol, Version
        6 (IPv6) Specification", RFC 1883, Xerox PARC, Ipsilon Networks,
        December 1995.

   [2] Hinden, R., and S. Deering, Editors, "IP Version 6 Addressing
        Architecture", RFC 1884, Ipsilon Networks, Xerox PARC, December
        1995.

AUTHORS' ADDRESSES

   Yakov Rekhter
   cisco Systems, Inc.
   470 Tasman Dr.
   San Jose, CA 95134

   Phone: (914) 528-0090
   EMail: yakov@cisco.com

   Tony Li
   cisco Systems, Inc.
   470 Tasman Dr.
   San Jose, CA 95134

   Phone: (408) 526-8186
   EMail: tli@cisco.com