Network Working Group   Y. Rekhter
Request for Comments: 1518   T.J. Watson Research Center, IBM Corp.
Category: Standards Track   T. Li
                                                          cisco Systems
                                                                Editors
                                                       September 1993

An Architecture for IP Address Allocation with CIDR

Status of this Memo

This RFC specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" for the standardization state and status of this protocol. Distribution of this memo is unlimited.

1. Introduction

This paper provides an architecture and a plan for allocating IP addresses in the Internet. This architecture and the plan are intended to play an important role in steering the Internet towards the Address Assignment and Aggregating Strategy outlined in [1].

The IP address space is a scarce shared resource that must be managed for the good of the community. The managers of this resource are acting as its custodians. They have a responsibility to the community to manage it for the common good.

2. 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 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 IP address allocation within the Internet. The first is the set of administrative requirements for obtaining and allocating IP 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.

The architecture and recommendations provided in this paper are intended for immediate deployment. This paper specifically does not address long-term research issues, such as complex policy-based routing requirements.

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 IP address allocation in the Internet. Topics covered include:

- Benefits of encoding some topological information in IP 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 IP addressing and routing components;

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

- Allocation of the IP addresses by the Internet Registry;

- Choice of the high-order portion of the IP 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 IP address allocation, both technical and administrative, that are not covered in this paper. Topics not covered or mentioned only superficially include:

- 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 IP address or a portion of the IP address space has been allocated);

- How a routing domain (especially a site) should organize its internal topology or allocate portions of its IP 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 IP addresses.

3. Background

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

IP partitions the routing problem into three parts:

- routing exchanges between end systems and routers (ARP),

- routing exchanges between routers in the same routing domain (interior routing), and,

- routing among routing domains (exterior routing).

4. IP Addresses and Routing

For the purposes of this paper, an IP prefix is an IP address and some indication of the leftmost contiguous significant bits within this address. Throughout this paper IP address prefixes will be expressed as <IP-address IP-mask> tuples, such that a bitwise logical AND operation on the IP-address and IP-mask components of a tuple yields the sequence of leftmost contiguous significant bits that form the IP address prefix. For example a tuple with the value <193.1.0.0 255.255.0.0> denotes an IP address prefix with 16 leftmost contiguous significant bits.

When determining an administrative policy for IP 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 IP addresses be assigned according to topological routing structures. However, 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.

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

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

Alternatively, should the routing domain draw IP addresses for all the hosts within the domain from a single IP address prefix, boundary routing information can be summarized into the single IP 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 IP 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 growth in the future to an Internet which has tens or 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, however, it should be possible to significantly constrain the volume and the complexity of routing information by taking advantage of the existing hierarchical interconnectivity, as discussed 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 small 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 abbreviate the reachability information for a large number of routing domains as a single prefix. This approach therefore can allow a great deal of hierarchical abbreviation 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 IP 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.

4.1 Efficiency versus Decentralized Control

If the Internet plans to support a decentralized address administration [4], then there is a balance that must be sought between the requirements on IP addresses for efficient routing and the need for decentralized address administration. A proposal described in [3] offers an example of how these two needs might be met.

The IP address prefix <198.0.0.0 254.0.0.0> provides for administrative decentralization. This prefix identifies part of the IP address space allocated for North America. The lower order part of that prefix allows allocation of IP addresses along topological boundaries in support of increased data abstraction. Clients within North America use parts of the IP address space that is underneath the IP address space of their service providers. Within a routing domain addresses for subnetworks and hosts are allocated from the unique IP prefix assigned to the domain.

5. IP Address Administration and Routing in the Internet

The basic Internet routing components are service providers (e.g., backbones, regional networks), and service subscribers (e.g., sites or campuses). These components are arranged hierarchically for the most part. A natural mapping from these components to IP routing components is that providers and subscribers 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. Such sites can exchange routing information with their provider via interior routing protocol route leaking or via an exterior routing protocol. For the purposes of this discussion, the choice is not significant. 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:

- at some part within a routing domain,
- at the leaf routing domain,
- at the transit routing domain (TRD), and
- at the continental boundaries.

A point within a routing domain corresponds to a subnetwork. 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.

The greatest burden in transmitting and operating on routing information is at the top of the routing hierarchy, where routing information tends to accumulate. In the Internet, for example, providers must manage the set of network numbers for all networks reachable through 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 network numbers for all attached providers and their associated networks.

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 IP address prefix directly from the IP address space allocated for North America, or from the IP 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.

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 IP address space that benefit the entire community.

5.1 Administration of IP addresses within a domain

If individual subnetworks take their IP addresses from a myriad of unrelated IP 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 IP 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 IP 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.

This situation is roughly analogous to the present dissemination of routing information in the Internet, where each domain may have non-contiguous network numbers assigned to it. The result of allowing subnetworks within a routing domain to take their IP addresses from unrelated IP address spaces is flat routing at the A/B/C class network level. The number of IP prefixes that leaf routing domains would advertise is on the order of the number of attached network numbers; the number of prefixes a provider's routing domain would advertise is approximately the number of network numbers 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 as the Internet grows this will quickly become intractable. A greater degree of hierarchical information reduction is necessary to allow continued growth in the Internet.

5.2 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 prefix from its provider's prefix 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 IP networks. Under the new allocation scheme, they might instead 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 subnetworks and routing domains implicit in the fact that they operate a common routing protocol and are under the control of a single administration. The routing domain administration subdivides the domain into subnetworks. The routing domain represents the only path between a subnetwork and the rest of the internetwork. It is reasonable that this relationship also extend to include a common IP addressing space. Thus, the subnetworks within the leaf routing domain should take their IP addresses from the prefix assigned to the leaf routing domain.

5.3 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 a TRD is a direct provider. Each case is discussed separately below.

5.3.1 Direct Service Providers

It is interesting to consider whether direct service providers' routing domains should use their IP address space for assigning IP 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.

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 backbones 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.

Are leaf routing domains willing to accept prefixes derived from the direct providers? In the supplier/consumer model, the direct provider is offering connectivity as the service, priced according to its costs of operation. This includes the "price" of obtaining service from one or more indirect providers (e.g., backbones). In general, indirect providers will want to handle as few address prefixes as possible to keep costs low. In the Internet environment, which does not operate as a typical marketplace, leaf routing domains must be sensitive to the resource constraints of the providers (both direct and indirect). The efficiencies gained in inter-domain routing clearly warrant the adoption of IP address prefixes derived from the IP address space of the providers.

The mechanics of this scenario are straightforward. Each direct provider is given a unique small set of IP address prefixes, from which its attached leaf routing domains can allocates slightly longer IP 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 IP address prefix <198.1.0.0 255.255.0.0>, NIST could use a unique IP prefix of <198.1.0.0 255.255.240.0>.

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. 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).

5.3.2 Indirect Providers (Backbones)

There does not appear to be a strong case for direct providers to take their address spaces from the the IP 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 IP 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 IP addresses derived from a backbone is inconsistent with the nature of the relationship.

5.4 Multi-homed Routing Domains

The discussions in Section 5.3 suggest methods for allocating IP 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 IP 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 IP 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.

One possible solution is for each multi-homed organization to obtain its IP address space independently from the providers to which it is attached. This allows each multi-homed organization to base its IP assignments on a single prefix, and to thereby summarize the set of all IP addresses reachable within that organization via a single prefix. The disadvantage of this approach is that since the IP 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.

A second possible approach would be for multi-homed organizations to be assigned a separate IP 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.

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 backbone. For example, lets suppose that the U.S. National Widget Manufacturers and Researchers have set up a U.S.-wide backbone, 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 backbone 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 backbone. However, since the widget backbone does not inform other general worldwide TRDs of what addresses it can reach (since the backbone 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.

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.

There are therefore a number of possible solutions to the problem of assigning IP 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 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 IP 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.

5.5 Private Links

The discussion up to this point concentrates on the relationship between IP 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 IP address allocation, and do not pose a problem for routing. 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 IP 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 5.4, 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 IP 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 IP 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" backbone, plus a number of private links to other leaf routing domains, can be treated as if it were single-homed to the backbone for the purpose of IP address allocation. We expect that this is also another way of dealing with multi-homed domains.

5.6 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 IP 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 IP addressing plan. This domain must be able to distinguish between the zero-homed routing domain's IP addresses and any other IP 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 IP addresses.

5.7 Continental aggregation

Another level of hierarchy may also be used in this addressing scheme to further reduce the amount of routing information necessary for inter-continental 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 subset of the address space. 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.

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 <194.0.0.0 254.0.0.0>, such that European routing also contains the longer prefixes <194.1.0.0 255.255.0.0> and <194.2.0.0 255.255.0.0>. 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 <194.0.0.0 255.0.0.0> into North American routing. Packets which are destined for 194.1.1.1 would traverse North American routing, but would encounter the North American router which performed the European aggregation. If the prefix <194.1.0.0 255.255.0.0> is unreachable, the router would drop the packet and send an ICMP Unreachable without using the trans-Atlantic link.

5.8 Transition Issues

Allocation of IP addresses based on connectivity to TRDs is important to allow scaling of inter-domain routing to an internet containing millions of routing domains. However, such address allocation based on topology implies that in order to maximize the efficiency in routing gained by such allocation, certain changes in topology may suggest a change of address.

Note that an address change need not happen immediately. A domain which has changed service providers may still advertise its prefix through its new service provider. Since upper levels in the routing hierarchy will perform routing based on the longest prefix, reachability is preserved, although the aggregation and scalability of the routing information has greatly diminished. Thus, a domain which does change its topology should change addresses as soon as convenient. The timing and mechanics of such changes must be the result of agreements between the old service provider, the new provider, and the domain.

This need to allow for change in addresses is a natural, inevitable consequence of routing data abstraction. The basic notion of routing data abstraction is that there is some correspondence between the address and where a system (i.e., a routing domain, subnetwork, or end system) is located. Thus if the system moves, in some cases the address will have to change. If it

were possible to change the connectivity between routing domains without changing the addresses, then it would clearly be necessary to keep track of the location of that routing domain on an individual basis.

In the short term, due to the rapid growth and increased commercialization of the Internet, it is possible that the topology may be relatively volatile. This implies that planning for address transition is very important. Fortunately, there are a number of steps which can be taken to help ease the effort required for address transition. A complete description of address transition issues is outside of the scope of this paper. However, a very brief outline of some transition issues is contained in this section.

Also note that the possible requirement to transition addresses based on changes in topology imply that it is valuable to anticipate the future topology changes before finalizing a plan for address allocation. For example, in the case of a routing domain which is initially single-homed, but which is expecting to become multi-homed in the future, it may be advantageous to assign IP addresses based on the anticipated future topology.

In general, it will not be practical to transition the IP addresses assigned to a routing domain in an instantaneous "change the address at midnight" manner. Instead, a gradual transition is required in which both the old and the new addresses will remain valid for a limited period of time. During the transition period, both the old and new addresses are accepted by the end systems in the routing domain, and both old and new addresses must result in correct routing of packets to the destination.

During the transition period, it is important that packets using the old address be forwarded correctly, even when the topology has changed. This is facilitated by the use of "longest match" inter-domain routing.

For example, suppose that the XYZ Corporation was previously connected only to the NorthSouthNet regional. The XYZ Corporation therefore went off to the NorthSouthNet administration and got an IP address prefix assignment based on the IP address prefix value assigned to the NorthSouthNet regional. However, for a variety of reasons, the XYZ Corporation decided to terminate its association with the NorthSouthNet, and instead connect directly to the NewCommercialNet public data network. Thus the XYZ Corporation now has a new address assignment under the IP address prefix assigned to the NewCommercialNet. The old address for the XYZ Corporation would seem to imply that traffic for the XYZ Corporation should be

routed to the NorthSouthNet, which no longer has any direct connection with XYZ Corporation.

If the old TRD (NorthSouthNet) and the new TRD (NewCommercialNet) are adjacent and cooperative, then this transition is easy to accomplish. In this case, packets routed to the XYZ Corporation using the old address assignment could be routed to the NorthSouthNet, which would directly forward them to the NewCommercialNet, which would in turn forward them to XYZ Corporation. In this case only NorthSouthNet and NewCommercialNet need be aware of the fact that the old address refers to a destination which is no longer directly attached to NorthSouthNet.

If the old TRD and the new TRD are not adjacent, then the situation is a bit more complex, but there are still several possible ways to forward traffic correctly.

If the old TRD and the new TRD are themselves connected by other cooperative transit routing domains, then these intermediate domains may agree to forward traffic for XYZ correctly. For example, suppose that NorthSouthNet and NewCommercialNet are not directly connected, but that they are both directly connected to the BBNet backbone. In this case, all three of NorthSouthNet, NewCommercialNet, and the BBNet backbone would need to maintain a special entry for XYZ corporation so that traffic to XYZ using the old address allocation would be forwarded via NewCommercialNet. However, other routing domains would not need to be aware of the new location for XYZ Corporation.

Suppose that the old TRD and the new TRD are separated by a non-cooperative routing domain, or by a long path of routing domains. In this case, the old TRD could encapsulate traffic to XYZ Corporation in order to deliver such packets to the correct backbone.

Also, those locations which do a significant amount of business with XYZ Corporation could have a specific entry in their routing tables added to ensure optimal routing of packets to XYZ. For example, suppose that another commercial backbone "OldCommercialNet" has a large number of customers which exchange traffic with XYZ Corporation, and that this third TRD is directly connected to both NorthSouthNet and NewCommercialNet. In this case OldCommercialNet will continue to have a single entry in its routing tables for other traffic destined for NorthSouthNet, but may choose to add one additional (more specific) entry to ensure that packets sent to XYZ Corporation's old address are routed correctly.

Whichever method is used to ease address transition, the goal is that knowledge relating XYZ to its old address that is held throughout the global internet would eventually be replaced with the new information. It is reasonable to expect this to take weeks or months and will be accomplished through the distributed directory system. Discussion of the directory, along with other address transition techniques such as automatically informing the source of a changed address, are outside the scope of this paper.

Another significant transition difficulty is the establishment of appropriate addressing authorities. In order not to delay the deployment of this addressing scheme, if no authority has been created at an appropriate level, a higher level authority may allocated addresses instead of the lower level authority. For example, suppose that the continental authority has been allocated a portion of the address space and that the service providers present on that continent are clear, but have not yet established their addressing authority. The continental authority may foresee (possibly with information from the provider) that the provider will eventually create an authority. The continental authority may then act on behalf of that provider until the provider is prepared to assume its addressing authority duties.

Finally, it is important to emphasize, that a change of addresses due to changes in topology is not mandated by this document. The continental level addressing hierarchy, as discussed in Section 5.7, is intended to handle the aggregation of reachability information in the cases where addresses do not directly reflect the connectivity between providers and subscribers.

5.9 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.

6. 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 IP addresses. As CIDR [1] is introduced in the Internet, it is therefore essential

to choose a hierarchical structure for IP addresses with great care.

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 IP address allocation, this again means that routing data abstraction must be encouraged.

In order for data abstraction to be possible, the assignment of IP 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 IP addresses assigned within a single routing domain use a single address prefix assigned to that domain. Specifically, this allows the set of all IP 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.

6.1 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 (Alternet, ANSnet, CIX, EBone, PSI, SprintLink);

- 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). However, the total number of such TRDs is expected to remain (for the foreseeable future) small enough to allow addressing of this set of TRDs via a flat address space. 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 IP 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 abbreviation. 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.

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

6.2 Recommendations for Multi-Homed Routing Domains

There are several possible ways that these multi-homed routing domains may be handled, as described in Section 5.4. 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.

6.3 Recommendations for the Administration of IP addresses

A companion document [3] provides recommendations for the administrations of IP addresses.

7. Acknowledgments

The authors would like to acknowledge the substantial contributions made by the authors of RFC 1237 [2], Richard Colella, Ella Gardner, and Ross Callon. The significant concepts (and a large portion of the text) in this document are taken directly from their work.

The authors would like to acknowledge the substantial contributions made by the members of the following two groups, the Federal Engineering Planning Group (FEPG) and the International Engineering Planning Group (IEPG). This document also reflects many concepts expressed at the IETF Addressing BOF which took place in Cambridge, MA in July 1992.

We would also like to thank Peter Ford (Los Alamos National Laboratory), Elise Gerich (MERIT), Steve Kent (BBN), Barry Leiner (ADS), Jon Postel (ISI), Bernhard Stockman (NORDUNET/SUNET), Claudio

Topolcic (CNRI), and Kannan Varadhan (OARnet) for their review and constructive comments.

8. References

[1] Fuller, V., Li, T., Yu, J., and K. Varadhan, "Supernetting: an Address Assignment and Aggregation Strategy", RFC 1338, BARRNet, cicso, Merit, OARnet, June 1992.

[2] Colella, R., Gardner, E, and R. Callon, "Guidelines for OSI NSAP Allocation in the Internet", RFC 1237, JuNIST, Mitre, DEC, July 1991.

[3] Gerich, E., "Guidelines for Management of IP Address Space", RFC 1466, Merit, May 1993.

[4] Cerf, V., "IAB Recommended Policy on Distributing Internet Identifier Assignment and IAB Recommended Policy Change to Internet "Connected" Status", RFC 1174, CNRI, August 1990.

9. Security Considerations

Security issues are not discussed in this memo.

10. Authors' Addresses

Yakov Rekhter
T.J. Watson Research Center, IBM Corporation
P.O. Box 218
Yorktown Heights, NY 10598

Phone: (914) 945-3896
EMail: yakov@watson.ibm.com

Tony Li
cisco Systems, Inc.
1525 O'Brien Drive
Menlo Park, CA 94025

EMail: tli@cisco.com