Network Working Group
Request for Comments: 1482

Mark Knopper
Steven J. Richardson
Merit/NSFNET
June 1993

Aggregation Support in the NSFNET Policy-Based Routing Database

Status of this memo

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

Abstract

This document describes plans for support of route aggregation, as specified in the descriptions of Classless Inter-Domain Routing (CIDR) [1] and the BGP-4 protocol [2], by the NSFNET Backbone Network Service. Mechanisms for exchange of route aggregates between the backbone service and regional/midlevel networks are specified. Additionally, the memo proposes the implementation of an Aggregate Registry which can be used by network service providers to share information about the use of aggregation. Finally, the operational impact of incorporating CIDR and aggregation is considered, including an analysis of how routing table size will be affected. This impact analysis will be used to modify the deployment plan, if necessary, to maximize operational stability.

1. Introduction

The Internet network service provider community and router vendors (as well as the IESG and various IETF working groups) have agreed that the time for deployment of route aggregation is upon us. This topic has been discussed in the BGP-D, NJM and ORAD working groups at several IETF meetings; it was a discussion topic of the NSFNET Regional Techs' Meetings in January and June, 1993; and it was also a topic of several meetings of the Federal Engineering Planning Group and Engineering and Operations Working Group of the Federal Network Council.

All have generally agreed that Summer, 1993 is the time to enable BGP-4 and CIDR aggregation. Each of the parties is responsible for its own aspect of CIDR implementation and practice. This memo describes Merit's plans for support of route aggregation on the NSFNET, and a proposal for implementing a database of aggregation information for use by network providers.

2. Aggregation Support by the Backbone Service

The NSFNET backbone service includes a Policy-Based Routing Database system which currently holds the set of network numbers that are accepted by the backbone service with a list of Autonomous System numbers from which announcements of these network numbers are expected. In order to implement CIDR, the database system will be modified to allow aggregation of routing information to be configured.

The NSFNET will (initially) not support de-aggregation on its outbound announcements. See section 2.3.

2.1 Current Configuration Capabilities

2.1.1 Inbound Announcements

An example of the way a network number is currently configured is as follows:

```
35      1:237   2:233   3:183   4:266   5:267  6:1225
```

This shows that network number 35 (ie. 35.0.0.0, a class A net number) is configured on the T3 backbone such that routing announcements are expected from up to 6 autonomous systems. The primary path is via AS 237, secondary is via AS 233, etc.

2.1.2 Outbound Announcements

Currently the NSFNET database has a list of AS's or network numbers for each neighbor AS that are announced by the backbone to that AS. These announcements are specified currently by "announcetoAS" statements--which implement policies submitted by midlevels to Merit--and then included in the ANSnet router configuration files. There are two forms of these statements. The first form uses the "norestrict" clause and indicates that all of the network numbers within each AS in the list should be announced to the neighbor midlevel AS. For example:

```
announcetoAS 42 norestrict ASlist 22 26 38 60 68
```

In this example, the NSFNET is configured to announce to neighboring midlevel AS 42, all networks in the routing table that were announced from AS's 22, 26, 38, 60 and 68.

If the "norestrict" keyword is changed to "restrict", this indicates that an explicit announce list of network numbers for the AS is specified in the configuration file. The NSFNET will only announce

network numbers that were announced by the AS's in the list, *AND* which appear in the "restrict list" of network numbers submitted separately by the midlevel.

For example,

announcetoAS 42 restrict ASlist 22

announce 192.135.237 <other info>

These statements mean that AS 42 only wishes to hear announcements from the backbone about the nets in AS 22 which are explicitly listed here (i.e., net 192.135.237).

It is also possible, when using the "restrict" keyword, to list specific "noannounce" lines. Those indicate that all of the networks listed in the routing table for the AS should be announced except those listed on the noannounce clauses. (There is also a "noannouncetoAS" statement[4].)

2.2 New Configuration Features for Aggregation

There will be three new capabilities for which the backbone service can be configured to support aggregation. The first two allow aggregates to be accepted and stored in the backbone routing tables based on announcements by the regional network (autonomous system or AS) peers. The third allows the announcement of aggregates to the AS neighbor peers. The following sections give examples of the three features.

We use the notation <net-IP prefix-length> to describe an aggregate. This refers to the IP prefix "net-IP", with a mask which has "prefix-length" 1's as counted from the high-order end. For example, <192.64.128 17> is equivalent to <192.64.128, 255.255.128.0> [5]. (The form using prefix-length rather than the mask is more compact.)

2.2.1 NSFNET accepts aggregates

In this case the regional peer router is CIDR-capable (i.e., runs BGP-4) and the announcement comes into the backbone as an IP address prefix.

To illustrate this in the spirit of sec. 2.1.1:

<192.64.128 17> 1:189 2:24 3:267

In this example, independent of the "class" of IP network number, an aggregate containing network addresses matching a pattern in which

the first 17 bits match the prefix 192.64.128 will be accepted in announcements to the NSFNET service. The primary path to destinations covered by the prefix is expected via AS 189, the secondary, via AS 24, etc.

2.2.2 NSFNET aggregates by proxy

The other method of incorporating CIDR aggregate announcements into the backbone routing tables is that of aggregation by proxy. In this case, the backbone is configured to perform aggregation on behalf of a peer AS which is not configured to announce the aggregate to the backbone (i.e., an AS which does not connect to the backbone via a CIDR-capable peer).

An example of this aggregation technique is:

```
proxy <192.64.128 17>     1:189  2:24  3:267
        if  <192.64.192 24>
        or  <192.64.129 24>
        or  <192.64.167 24>
```

(Note: the syntax used in this document is arbitrary and is only used to illustrate the method. The syntax to be used in actual routing requests is to be determined.)

In this example, the aggregate <192.64.128 17> will be stored and propagated within the backbone as an aggregate under a set of conditions. Initially, the GateD support will allow an "OR" list of conditions such that if one of the aggregates in the list matches the proxy aggregate will be stored[6]. For the case above, this means that, if any of the CIDR aggregates:

```
<192.64.192 24>
<192.64.129 24>
<192.64.167 24>
```

(which--under the current, class-based IP address system--are equivalent to the class C net numbers 192.64.192, 192.64.129, or 192.64.167, respectively) is heard, the backbone router will act as though it heard the announcement of the single CIDR aggregate <192.64.128 17>.

2.2.3 NSFNET announces aggregates

The functionality of the current system, as outlined in sec. 2.1.2, above, will continue to exist once CIDR is implemented. The "norestrict" function (or its equivalent in the new software) will specify that all network reachability information received from a set

of Autonomous Systems, including any aggregates, will be announced. It should also be possible to use to the equivalents of the "restrict" keyword and the "announce" (or "noannounce") statement in order to limit the announcements of the aggregations within an AS to any desired subset.

2.3 Specifically Unsupported Capabilities, Limits of Initial Deployment

There are some aspects of aggregation which will specifically not be supported in the initial deployment of CIDR capabilities on the NSFNET backbone. In particular, when the NSFNET service announces routes to midlevel peers, de-aggregation will not be performed [3]. Therefore, a peer which needs to receive full routing information should run a protocol which supports CIDR (initially, BGP-4; later, IDRP). Peer networks using default routing will be able to reach networks that are part of aggregated routing information across the backbone (as in section 6.4 of [3]).

3. CIDR Aggregate Registry

In discussions with network service providers, it has become apparent that there is a great need for sharing of aggregate information; this is necessary to fulfill the coordination referred to in sec. 2.3. Beyond the need to implement CIDR aggregation facilities in the NSFNET Policy-Based Routing Database (as described in section 2), there is a clear need to have a separate database which will allow aggregate information from any Autonomous System to be stored and made available for easy electronic retrieval. This information can be used for routing coordination and policy configuration in the larger, non-NSFNET-centric, inter-domain context.

One of the expected uses of such a database is to help determine, as CIDR matures, the granularity of aggregation of network reachability information with respect to policy. The useful scope of aggregation is the subject of much discussion[5][7], and will be influenced by such considerations as how network number allocation has been handled, and whether the network provider has renumbered its client networks to conform to CIDR aggregation boundaries. Rules and issues regarding network number allocation with CIDR are discussed in [8] and [7].

In order further these goals, Merit proposes to implement a "CIDR Aggregate Registry" to provide sharing of aggregate information for the Internet inter-domain routing community. Initially, this will be a simple database without much structure. It is not intended to hold only aggregates which are announced or accepted by the NSFNET service; rather, it should be a community registry that all will be invited to use and make use of.

The Aggregate Registry will consist of a list of aggregate announcement statements. Each statement consists of four types of information, along with contact information:

1) CIDR Aggregate: The aggregate identifier, consisting of a network number prefix and the prefix length. For example, <192.29.128 16>.

2) Home AS: The source AS number for the aggregate. That is, the AS number of the network service provider that initially aggregates the network reachability information into the aggregate for announcement to its neighbors.

3a) Announcing AS: An AS number that announces this aggregate to its neighbor AS's.

3b) Neighbor AS list: A list of neighbor AS's to whom the aggregate will be announced by the AS named in 3a.

4) Contact information: eg. e-mail address and name or NIC handle of the administrative and technical contacts for the source AS.

Thus, a given aggregate is listed once as announced by its source AS. It may then be listed once again per transit AS which announces the aggregate downstream to its neighbors. For example, the CIDR aggregate <199.29.128 16> could be listed as:

CIDR aggregate (prefix-length)	home AS	ann AS	neighbor AS list	contacts
<199.29.128 16>	100	100	200 201 690	fred@nowhere.net
<199.29.128 16>	100	690	266 267 1225...	<contact info>
<199.29.128 16>	100	200	297 372	<contact info>
<199.29.128 16>	100	201	771 1262	<contact info>

Note: This can be represented using the syntax used for objects in the RIPE-81 paper[9].

Here, AS 100 (the source AS) performs any aggregation and announces the CIDR aggregate <199.29.128 16> to neighbor ASs 200, 201, and 690. In turn, AS 200 announces this same aggregate to its neighbor ASs 297 and 372; further lines show announcements of the given aggregate by AS 690 and AS 201.

Note that this registry reflects both the simple list of aggregates that are supported by the union of network providers, as well as information on inter-domain topology for the Internet. Merit will implement procedures for registering any network provider's

aggregates in the Registry; for those CIDR aggregates carried over the NSFNET backbone, Merit will implement procedures for integrating this Registry with the process of updating the aggregate routing announcements. Requests to update the information will be handled via e-mail or on-line registration tools.

4. Effects of CIDR on Operational Aspects of the Internet

The introduction of CIDR will clearly necessitate various changes beyond the introduction of new router software. In particular, Merit and other network service providers will have to adjust tools, reports, and procedures as CIDR is implemented and evolved, and these changes will have to be coordinated in order to ensure a smooth transition to the CIDR-capable Internet.

While this document is by no means exhaustive, some of the areas affected are discussed briefly below; what is intended is to foster an awareness of some these changes, so as to initiate thinking about and planning for this transition. While it is obvious that CIDR and policy routing imply greater coordination of many operational matters, it is not clear how profoundly this will affect the day-to-day running of the Internet.

(Note: Aspects of the actual phased deployement of CIDR are covered in [3] and [10].)

4.1 NSFNET Configuration Files and Reports; Neighbor AS Configurations

The addition of CIDR capability to the NSFNET Policy-Based Routing Database, as outlined in sec. 2, will require the updating of at least the following reports which are currently produced by Merit (and available via anonymous FTP from nic.merit.edu):

ans_core.now as-site.now country.now net-comp.now net-net.now
net-ter.now non-us.now

Any tools which access this information, such as the various clients or scripts released by Merit or developed by others, will have to be changed.

However, the most striking change will be in the transition from rcp_routed to GateD; it is very different in important particulars, and follows different conceptual principles [11].

Network providers which develop any part of their configuration files from parsing the NSFNET configuration files or reports *MUST* plan for these changes in order to help themselves and the Internet community achieve a smooth transition to CIDR.

4.2 Routing and Administrative Policies

In this document, Merit has stated its commitment to supporting CIDR through both changing policies related to administering the NSFNET and developing a CIDR Aggregate Registry for the broader Internet community.

In addition to these changes, here are some of the other policies, administrative and routing, which must to be coodinated in order to achieve optimum benefits of CIDR:

- policies of the InterNIC and of network service providers in assigning (CIDR) IP nets and blocks, as mentioned above;

- policies of the various ASs in coordination of transit and other routing policies;

- policies of registration of new networks, from the InterNIC or network provider, through the CIDR Aggregate Registry, etc.;

- policies related to coordination of routing changes;

- coordination of routing policies, in general, to avoid new classes of routing problems due to new methods of routing.

4.3 Realtime Issues

Issues which have not been examined in detail are:

- debugging of routing/connectivity problems;

- stability and other properties of routing under various scenarios of CIDR configuration and network topology;

- explicit specification of routing decision algorithms to avoid routing anomalies;

- increased network load due to packets traversing an AS, such as the NSFNET backbone, before being discarded due to addressing a "hole" in a CIDR aggregate.

4.4 Estimate of Reductions in Routing Tables

An argument in favor of the implementation CIDR is the effect which it should have upon the NSFNET and other routing tables [1] [5]. The burning question is: What is the magnitude of this effect? In view of the various issues to be dealt with, this is an important consideration.

In terms of the immediate savings in reduction of the NSFNET backbone routing tables, if a set of aggregates were done all at once, a recent calculation--which might be characterized as an optimistic estimate using a pessimistic algorithm (it looks for the longest continuous block of addresses announced to the NSFNET backbone)--yields [12]:

```
 861 size  2 saving  861 announcements
 286 size  4 saving  858 announcements
 117 size  8 saving  819 announcements
  67 size 16 saving 1005 announcements
  13 size 32 saving  403 announcements
   3 size 64 saving  189 announcements
1347 total   saving 4135 announcements of 12348 (33%).
```

Here, the first column represents the number of CIDR aggregates of the given "size," and shows the corresponding reduction in net announcements due to the adoption of this aggregate. (A CIDR aggregate of "size <n>" is one which encompasses <n> class A, B, or C networks; the 67 "size 16" CIDR aggregates actually combine announcements for 16 separate networks into a single net aggregate.) It is unclear, at this time, whether or not the true savings would be of this magnitude, but the extended report provides a basis for discussion [12].

The other aspect of impact upon the routing tables, the reduction in the rate of growth (and the concomitant slowing of the rate of exhaustion of IP address space), is an entirely different matter. Simple calculations related to the rate of class B address space exhaustion indicate that CIDR-conformant policies of the InterNIC with respect to address assignment is helping [1].

Clearly, more detailed analysis is desirable in order to better understand the realistic gains of the CIDR deployment process, both initially and in the longer term.

5. Conclusions and Next Steps

Implementation of CIDR is underway, but there is still a fair amount of planning and discussion that is needed for a successful transition. Merit is proposing specific functions for CIDR aggregation that will be supported by the NSFNET, as well as a CIDR Aggregate Registry that can serve as the basis for inter-domain routing coordination.

The Aggregate Registry will allow a set of tools to be developed that can facilitate the design of aggregation policy. A query tool to allow lookup of aggregation information for a given network or

aggregate would be very useful. Additional database functionality will also be desired for more powerful queries. It is specifically a goal to work with RIPE to make sure that the Merit and RIPE database approaches are compatible and allow interworking of tools. An AS topology database would be most useful in routing policy determination and coordination as well.

In addition to these areas, many other issues require further work in order to develop the operational framework necessary for the successful use of CIDR on the Internet. It is critical that the deployment of CIDR and related tools to preserve address and routing table space must not compromise the operational stability of the NSFNET and the wider Internet.

6. Security Considerations

Security issues are not discussed in this document.

7. Acknowledgements

The authors would like to acknowledge the following persons, whose comments and discussions have helped to shape this document:

Dennis Ferguson, Advanced Network and Services, Inc.
Jeffrey Honig, Cornell University
William Manning, Rice University/SESQUINET
The Merit Internet Engineering and Network Management Systems groups.

8. Authors' Addresses

Knopper, Mark A.
Merit Network, Inc.
1071 Beal Ave.
Ann Arbor, MI 48109-2103

e-mail: mak@merit.edu
phone: (313) 763-6061
fax: (313) 747-3745

Richardson, Steven J.
Merit Network, Inc.
1071 Beal Ave.
Ann Arbor, MI 48109-2103

e-mail: sjr@merit.edu
phone: (313) 747-4813
fax: (313) 747-3745

9. References

[1] Fuller, V., Li, T., Yu, J., and Varadhan, K., "Supernetting: an Address Assignment and Aggregation Strategy", RFC1338, Update, Work in Progress, June 1992.

[2] Rekhter, Y., and Li, T., "A Border Gateway Protocol 4", Work In Progress, April 1993.

[3] Topolcic, C., "Notes of BGP-4/CIDR Coordination Meeting of 11 March 93", Work in Progress, March 1993.

[4] Villamizer, C., in a document describing rcp_routed.conf options and syntax, May, 1993.

[5] Syntax used in Ford, P., Rekhter, Y., Braun, H-W., "Improving the Routing and Addressing of IP", IEEE Network, pp. 10-15, May 1993.

[6] Ferguson, D., private correspondence, March, 1993.

[7] Rekhter, Y., and Li, T., "An Architecture for IP Address Allocation with CIDR", Work in Progress, February, 1993.

[8] Gerich, E., "Guidelines for Management of IP Address Space", RFC1466, May 1993.

[9] Bates, T., Jouanigot, J-M., Karrenberg, D., Lothberg, P., and Terpstra, M., "Representation of IP Routing Policies in the RIPE Database" (ripe-81), Work in Progress, February, 1993.

[10] Rekhter, Y., and Topolcic, C., "Exchanging Routing Information Across Provider/Subscriber Boundaries in the CIDR Environment", Work in Progress, April 1993.

[11] Fedor, M., Honig, J., Coltun, R., Ferguson, D., "gated-config(5)" manpage, from the "gated-R3_0Beta_2" distribution, 7 October 1992.

[12] Johnson, D., analysis available via anonymous FTP from merit.edu:/pub/nsfnet/cidr/auto-aggregates, June 1993.

[13] Topolcic, C., "Schedule for IP Address Space Management Guidelines", RFC1367, October, 1993.