                    The NSFNET Routing Architecture

Status of this Memo

   This document describes the routing architecture for the NSFNET
   centered around the new NSFNET Backbone, with specific emphasis on
   the interface between the backbone and its attached networks.
   Distribution of this memo is unlimited.

Introduction

   This document describes the routing architecture for the NSFNET
   centered around the new NSFNET Backbone, with specific emphasis on
   the interface between the backbone and its attached networks. It
   reflects and augments thoughts described in [1], discussions during
   the Internet Engineering Task Force meeting at the San Diego
   Supercomputing Center in March 1988, discussions on mailing lists,
   especially on a backbone/regional network working group mailing list,
   and a final discussion held at the IBM T.J. Watson Research Center in
   Yorktown, NY, on the 21st of March 1988. The Yorktown meeting was
   attended by Hans-Werner Braun (Merit), Scott Brim (Cornell
   University), Mark Fedor (NYSERNet), Jeff Honig (Cornell University),
   and Jacob Rekhter (IBM). Thanks also to: Milo Medin (NASA), John Moy
   (Proteon) and Greg Satz (Cisco) for discussing this document by email
   and/or phone.

   Understanding of [1] is highly recommended prior to reading this
   document.

1. Routing Overview

   The new NSFNET backbone forms the core of the overall NSFNET, which
   connects to regional networks (or regional backbones) as well as to
   peer networks (other backbones like the NASA Science Network or the
   ARPANET). The NSFNET core uses a SPF based internal routing
   protocol, adapted from the IS-IS protocol submitted by ANSI for
   standardization to the ISO. The ANSI IS-IS protocol is based upon
   work done at Digital Equipment Corporation. Its adaptation to the
   Internet environment requires additional definitions, most notably to
   the addressing structure, which will be described in a later
   document. This adaptation was largely done by Jacob Rekhter of IBM
   Research in Yorktown, NY. The RCP/PSP routing architecture was
   largely implemented by Rick Boivie and his colleagues at IBM TCS in

Braun   [Page 1]

Milford, CT. The adaptation of EGP to the NSS routing code and the
new requirements was done jointly by Jeff Honig (who spent about a
week to work on this at IBM Research) and Jacob Rekhter. Jeff is
integrating the changes done for the new EGP requirements into the
"gated" distributions.

The IGP derives routing tables from Internet address information.
This information is flooded throughout the NSFNET core, and the
individual NSS nodes create or update their routing information after
running the SPF algorithm over the flooded information. A detailed
description of the NSFNET backbone IGP will be documented in a future
document.

The routing interface between the NSFNET core and regional backbones
as well as peer networks utilizes the Exterior Gateway Protocol
(EGP). The EGP/IGP consistency and integrity at the interface points
is ensured by filtering mechanisms according to individual nodal
routing policy data bases [1]. EGP is selected as the routing
interface of choice between the NSFNET backbone and its regional
attachments due to its widespread implementation as well its ability
to utilize autonomous system designators and to allow for effective
firewalls between systems. In the longer run the hope is to replace
the EGP interface with a new inter Autonomous System protocol. Such a
new protocol should also allow to move the filtering of network
numbers or Autonomous Network number groups to the regional gateways
in order for the regional gateways to decide as to what routing
information they wish to receive.

A general model is to ensure consistent routing information between
peer networks. This means that, e.g., the NSFNET core will have the
same sets of Internet network numbers in its routing tables as are
present in the ARPANET core. However, the redistribution of this
routing information is tightly controlled and based on Autonomous
System numbers. For example, ARPANET routes with the ARPANET
Autonomous System number will not be redistributed into regional or
other peer networks. If an NSFNET internal path exists to such a
network known to the ARPANET it may be redistributed into regional
networks, subject to further policy verification. Generally it may be
necessary to have different trust models for peer and subordinate
networks, while giving a greater level of trust to peer networks.

The described use of EGP, which is further elaborated on in [1]
requires bidirectional translation of network information between the
IGP in use and EGP.

2. Conclusions reached during the discussions

The following conclusions were reached during the meeting and in

subsequent discussions:

No DDN-only routes (ARPANET/MILNET) shall be announced into the
regional backbones. This is a specific case of the ability to
suppress information from specific Autonomous Systems, as
described later.

Regional backbones are required to use an unique Autonomous System
number. Announcements from non-sanctioned autonomous systems,
relative to a particular site, will not be believed and will
instead trigger an alarm to the Network Operations Center.

Regional backbone attachments must not require routes to local
subnets. This means that the locally attached network needs to
use a flat space, without subnet bits, at least from the NSS point
of view. The reason for this is that the EGP information
exchanged between the regional gateway and the NSS cannot include
subnet information. Therefore the NSS has no knowledge of remote
subnets. The safest way to get around this limitation is to use a
non-subnetted network (like a separate Class-C network) at the
interface between a regional backbone and the NSFNET backbone.
The other way is to use Proxy-ARP while having just the NSS think
that the network is not subnetted. In the latter case care must be
taken so that the E-PSP uses the proper local IP broadcast
address.

Routing information received by the NSS from regional gateways
will be verified on both network number and autonomous system
number.

Metric reconstitution is done on a per-network basis. The NSS
will construct the fixed metric it will use for a given network
number from its internal data base. Network metrics given to the
NSS via EGP will be ignored. The metrics used are a result of an
ordered list of preferred paths as supplied by the regional
backbones and the attached campuses. This metric is of relevance
only to the NSFNET core itself. The mechanisms are further
explained in [1].

Global metric reconstitution by Autonomous System numbers is
necessary in specific cases, such as peer networks. An example is
that ARPANET routes will be reconstituted to a global metric, as
determined by the NSS.

EGP announcements into regional networks will use a fixed metric.
The metric used shall be "128." The 128-metric is somewhat
arbitrarily chosen to be high enough so that a regional backbone
will get a metric high enough from the NSFNET Core AS to allow a

comparison against other (most likely internal) routes. "128" is
also consistent with [2].

Peer network routes (e.g., ARPANET routes) are propagated through
the NSS structure.

No DEFAULT routing information is distributed within the NSFNET
backbone, as the NSFNET core has the combined routing knowledge of
the attached regional and peer networks.

We do not expect the requirement for damping of routing update
frequencies, at least initially. The frequency of net up/down
changes combined with the available bandwidth and CPU capacity do
not let the frequency of SPF floodings appear as being a major
problem. Simple metric changes as heard by a NSS via EGP will not
trigger updates.

An allowed list of Source Autonomous System information will be
used to convert from the IGP to EGP, on a Destination Autonomous
System number basis, to allow for specific exclusion of definable
remote Autonomous System information.

EGP must only announce networks for which the preferred path is
via the IGP. This means in particular that the EGP peer will
never announce via EGP what it learned via EGP on the same
interface, not even if the information was received from a third
EGP peer. This will avoid the back-distribution of information
learned via that same interface. The EGP peers of regional
gateways must only announce information belonging to their own
Autonomous System.

EGP will be used in interior mode only.

The regional backbones are responsible for generating DEFAULT
routing information at their option. One possibility is to
generate an IGP default on a peer base as long as the NSS EGP
connection is working. The EGP information will not include a
special indication for DEFAULT.

It is highly desirable to have direct peer-peer connections, to
ease the implementation of a consistent routing data base.

A single Autonomous System number may not be used with two E-PSPs
at the same time as long as the two E-PSP's belong to the same
NSS. Otherwise the same Autonomous System number can be used from
multiple points of attachment to the backbone and therefore can
talk to more than one E-PSP. However, this may result in
suboptimal routing unless multiple announcements are properly

engineered according to [1].

The administrator of the regional networks should be warned that
improper routing implementations within the region may create
suboptimal regional routing by using this restriction if no care
is taken in that:

Only networks belonging to their own Autonomous System get
preferred over NSFNET backbone paths; this may extend to a
larger virtual Autonomous System if backdoor paths are
effectively implemented.

IGP implementations should not echo back routing information
heard via the same path.

If two regional networks decide to implement a backdoor
connection between themselves, then the backdoor must have a
firewall in so that information about their own Autonomous
System cannot flow in from the other Autonomous System. That
is, a regional network must not allow information about
networks that are interior to its Autonomous System to enter
via exterior routes. Likewise, if a regional network is
connected to the NSFNET via two NSS connections, the NSS cannot
send back information about the Autonomous System into the
Autonomous System where it originated. The end effect is that
partitions within an Autonomous System will not be healed by
using the NSS system. In addition, if three or more regionals
connect to each other via multiple back-door paths, it is
imperative that all back-door paths have firewalls that ensure
that the above restrictions are imposed. These actions are
necessary to prevent routing loops that involve the NSS system.
Furthermore routing information should only be accepted from
another regional backbone via backdoor paths for networks which
are positively desired to be reached via this same backdoor
path.

3. EGP requirements for attached gateways

The following EGP requirements are necessary for attached gateways;
they may require changes in existing vendor products:

IGP to EGP routing exchanges need to be bidirectional. This
feature should be selectable by the gateway administrator, and by
default be configured OFF.

The metric used when translating from EGP to IGP should be
configurable.

It must be possible for IGP information to override EGP
information, so that the internal paths are preferred over
external paths. Overriding EGP information on an absolute basis,
where an external path would never be used as long as there is an
internal one, is acceptable.

The ability to do route filtering in the regional gateways on a
per net basis is highly desirable to allow the regional gateways
to do a further selection as to what routes they would want to
redistribute into their network.

The existence of an EGP connection should optionally lead to the
generation of a DEFAULT announcement for propagation via the IGP.
The DEFAULT metric should be independently configurable.

EGP routes with a metric of "128" should be acceptable. In most
cases the regional backbone should ignore the EGP metric.

The regional gateways must only announce networks known to their
own Autonomous System. At the very least they must not
redistribute routing information via EGP for routes previously
learned via EGP.

It would be beneficial if the regional IGPs would tag routes as
being EGP derived.

If the EGP peer (e.g., a NSS) terminates the EGP exchange the
previously learned routes should expire in a timely fashion.

4. References

   [1]   Rekhter, J., "EGP and Policy Based Routing in the New NSFNET
         Backbone", T.J. Watson Research Center, IBM Corporation, March
         1988. Also as RFC 1092, February 1989.

   [2]   Mills, D., "Autonomous Confederations", RFC 975, M/A-COM
         Linkabit, February 1986.

   [3]   Mills, D., "Exterior Gateway Formal Specification", RFC 904,
         M/A-COM Linkabit, April 1984.

   [4] "Exterior Gateway Protocol, Version 3, Revisions and Extensions,"
         Working Notes of the IETF WG on EGP, Marianne L. Gardner and
         Mike Karels, February 1988.

   [5]   "Management and Operation of the NSFNET Backbone Network,"
         proposal to the National Science Foundation, Merit Computer
         Network, August 1987.

5. Appendix

   The following are extensions implemented for the "gated" EGP
   implementation, as designed by Jeff Honig of the Cornell University
   Theory Center. These extensions are still in the design stage and
   may be changed over time. They are included here as an
   implementation example.

   Changes to egpneighbor clause:

   egpneighbor <address>   metricin <metric>
                            egpmetricout <egpmetric>
                            ASin <as>
                            ASout <as>
                            nogendefault
                            acceptdefault
                            defaultout <egpmetric>
                            validate

   metricin <metric>

        If specified, the metric of all nets received from this
        neighbor are set to <metric>.

   egpmetricout <egpmetric>

        If specified, the metric of all nets sent to this neighbor,
        except default, are set to <egpmetric>.

   ASin <as>

        If specified, EGP packets received from this neighbor must
        specify this AS number of an EGP error packet is generated.
        The AS number is only checked at neighbor acquisition time.

   ASout <as>

        If specified, this AS number is used on all EGP packets sent
        to thiqs neighbor

   nogendefault

        If specified, this neighbor is not considered when
        generating a gateway default.

   acceptdefault

        If specified, the default will be accepted from this

        neighbor, otherwise it will be explicitly ignored.

    defaultout <egpmetric>

        If specified, the internally generated default is send to
        this neighbor in EGP updates. Default learned from other
        gateways is not propogated.

    validate

        If specifed, all nets learned from this EGP neighbor must
        have a corresponding 'validAS' clause or they will be
        ignored.

    Addition of a validAS clause:

    validAS <net> AS <as> metric <metric>

        This clause specifies which AS a network may be learned from and
        what internal metric to use when the net is learned. The
        specifies the 'validate' option. Note that more than one may be
        learned from more than one AS.

    Addition of sendAS and donotsendAS clauses:

        These clauses control the announcement of exterior (currently only
        EGP) routes. Normally, exterior routes are not considered for
        announcement. When the 'sendAS' or 'donotsendAS' clauses are
        used, the announce/donotannounce, egpnetsreachable and other
        restrictions still apply. The 'sendAS' and 'donotsendAS' clauses
        are mutually exclusive by autonomous system.

    sendAS <as0> ASlist <as1> <as2> ...

        This clause specifies that only nets learned from as1, as2, ...
        may be propogated to as0.

    donotsendAS <as0> ASlist <as1> <as2> ...

        This clause specifies that nets learned from as1, as2, ... may
        not be propogated to <as0>, all other nets are propogated.

    An example of a "/etc/gated.conf" file could include the following:

    #
    RIP supplier
    #
    autonomousystem (regional AS)

```
     #
     egpneighbor (NSS address) ASin (NSS AS) nogendefault
     metricin (metric)
     #
     sendAS (NSS AS) ASlist (regional AS)
     #
```

     Where:

```
          Regional AS    Is the AS number of the regional network
          NSS address    Is the IP address of the local NSS
          NSS AS         Is the AS number the NSFNET backbone
          Metric         Is the gated internal (time delay) metric that
                         EGP learned routes should have.  This is the
                         metric used on output after conversion to a RIP
                         metric.  Some values are:
```

```
                              HELLO   RIP
                              100     1
                              148     2
                              219     3
                              325     4
                              481     5
```

Author's Address:

     Hans-Werner Braun
     University of Michigan
     Computing Center
     1075 Beal Avenue
     Ann Arbor, MI 48109

     Phone: (313) 763-4897

     Email: HWB@MCR.UMICH.EDU