                      Autonomous Confederations

Status of This Memo

   This RFC proposes certain enhancements of the Exterior Gateway
   Protocol (EGP) to support a simple, multiple-level routing capability
   while preserving the robustness features of the current EGP model.
   It requests discussion and suggestions for improvements.
   Distribution of this memo is unlimited.

Overview

   The enhancements, which do not require retrofits in existing
   implementations in order to interoperate with enhanced
   implementations, in effect generalize the concept of core system to
   include multiple communities of autonomous systems, called autonomous
   confederations. Autonomous confederations maintain a higher degree of
   mutual trust than that assumed between autonomous systems in general,
   including reasonable protection against routing loops between the
   member systems, but allow the routing restrictions of the current EGP
   model to be relaxed.

   The enhancements involve the "hop count" or distance field of the EGP
   Update message, the interpretation of which is not covered by the
   current EGP model. This field is given a special interpretation
   within each autonomous confederation to support up to three levels of
   routing, one within the autonomous system, a second within the
   autonomous confederation and an optional third within the universe of
   confederations.

1.   Introduction and Background

   The historical development of Internet exterior-gateway routing
   algorithms began with a rather rigid and restricted topological model
   which emphasized robustness and stability at the expense of routing
   dynamics and flexibility. Evolution of robust and dynamic routing
   algorithms has since proved extraordinarily difficult, probably due
   more to varying perceptions of service requirements than to
   engineering problems.

   The original exterior-gateway model suggested in RFC-827 [1] and
   subsequently refined in RFC-888 [2] severely restricted the Internet
   topology essentially to a tree structure with root represented by the
   BBN-developed "core" gateway system. The most important
   characteristic of the model was that debilitating resource-consuming
   routing loops between clusters of gateways (called autonomous

systems) could not occur in a tree-structured topology. However, the
administrative and enforcement difficulties involved, not to mention
the performance liabilities, made widespread implementation
impractical.

1.1. The Exterior Gateway Protocol

   Requirements for near-term interoperability between the BBN core
   gateways and the remainder of the gateway population implemented
   by other organizations required that an interim protocol be
   developed with the capability of exchanging reachability
   information, but not necessarily the capability to function as a
   true routing algorithm. This protocol is called the Exterior
   Gateway Protocol (EGP) and is documented in RFC-904 [3].

   EGP was not designed as a routing algorithm, since no agreement
   could be reached on a trusted, common metric. However, EGP was
   designed to provide high-quality reachability information, both
   about neighbor gateways and about routes to non-neighbor gateways.
   At the present state of development, dynamic routes are computed
   only by the core system and provided to non-core gateways using
   EGP only as an interface mechanism. Non-core gateways can provide
   routes to the core system and even to other non-core gateways, but
   cannot pass on "third-party" routes computed using data received
   from other gateways.

   As operational experience with EGP has accumulated, it has become
   clear that a more decentralized dynamic routing capability is
   needed in order to avoid resource-consuming suboptimal routes. In
   addition, there has long been resistance to the a-priori
   assumption of a single core system, with implications of
   suboptimal performance, administrative problems, impossible
   enforcement and possible subversion. Whether or not this
   resistance is real or justified, the important technical question
   remains whether a more dynamic, distributed approach is possible
   without significantly diluting stability and robustness.

   This document proposes certain enhancements of EGP which
   generalize the concept of core system to include multiple
   communities of autonomous systems, called autonomous
   confederations. Autonomous confederations maintain a higher
   degree of mutual trust than that assumed between autonomous
   systems in general, including reasonable protection against
   routing loops between the member systems. The enhancements
   involve the "hop count" or distance field of the EGP Update

message, which is given a special interpretation as described
later. Note that the interpretation of this field is not
specified in RFC-904, but is left as a matter for further study.

The interpretation of the distance field involves three levels of
metrics, in which the lowest level is available to the interior
gateway protocol (IGP) of the autonomous system itself to extend
the interior routes to the autonomous system boundary. The next
higher level selects preferred routes within the autonomous system
to those outside, while the third and highest selects preferred
routes within the autonomous confederation to those outside.

The proposed model is believed compatible with the current
specifications and practices used in the Internet. In fact, the
entire present conglomeration of autonomous systems, including the
core system, can be represented as a single autonomous
confederation, with new confederations being formed from existing
and new systems as necessary.

1.2. Routing Restrictions

It was the intent in RFC-904 that the stipulated routing
restrictions superceded all previous documents, including RFC-827
and RFC-888. The notion that a non-core gateway must not pass on
third-party information was suggested in planning meetings that
occured after the previous documents had been published and before
RFC-904 was finalized. This effectively obsoletes prior notions
of "stub" or any other asymmetry other than the third-party rule.

Thus, the only restrictions placed on a non-core gateway is that
in its EGP messages (a) a gateway can be listed only if it belongs
to the same autonomous system (internal neighbor) and (b) a net
can be listed only if it is reachable via gateways belonging to
that system. There are no other restrictions, overt or implied.
The specification does not address the design of the core system
or its gateways.

The restrictions imply that, to insure full connectivity, every
non-core gateway must run EGP with a core gateway. Since the
present core-gateway implementation disallows other gateways on
EGP-neighbor paths, this further implies that every non-core
gateway must share a net in common with at least one core gateway.

Note that there is no a-priori prohibition on using EGP as an IGP,
or even on using EGP with a gateway of another non-core system,

        providing that the third-party rule is observed. If a gateway in
        each system ran EGP with a gateway in every other system, the
        notion of core system would be unneccessary and superflous.

        At one time during the evolution of the EGP model a strict
        hierarchical topology (tree structure) of autonomous systems was
        required, but this is not the case now. At one time it was
        forbidden for two nets to be connected by gateways of two or more
        systems, but this is not the case now. Autonomous systems are
        sets of gateways, not nets or hosts, so that a given net or host
        can be reachable via more than one system; however, every gateway
        belongs to exactly one system.

    1.3. Examples and Problems

        Consider the common case of two local-area nets A and B connected
        to the ARPANET by gateways of different systems. Now assume A and
        B are connected to each other by a gateway A-B belonging to the
        same system as the A-ARPANET gateway, which could then list itself
        and both the A and B nets in EGP messages sent to any other
        gateway, since both are now reachable in its system. However, the
        B-ARPANET gateway could list itself and only the B net, since the
        A-B gateway is not in its system.

        In principle, we could assume the existence of a second gateway
        B-A belonging to the same system as the B-ARPANET gateway, which
        would entitle it to list the A net as well; however, it may be
        easier for both systems to sign a treaty and consider the A-B
        gateway under joint administration. The implementation of the
        treaty may not be trivial, however, since the joint gateway must
        appear to other gateways as two distinct gateways, each with its
        own autonomous-system number.

        Another case occurs when for some reason or other a system has no
        path to a core gateway other than via another non-core system.
        Consider a third local-are net C, together with gateway C-A
        belonging to a system other than the A-ARPANET and B-ARPANET
        gateways. According to the restrictions above, gateway C-A could
        list net C in EGP messages sent to A-ARPANET, while A-ARPANET
        could list ARPANET in messages sent to C-A, but not other nets
        which it may learn about from the core. Thus, gateway C-A cannot
        acquire full routing information unless it runs EGP directly with
        a core gateway.

2.   Autonomous Systems and Confederations

   The second example above illustrates the need for a mechanism in
   which arbitrary routing information can be exchanged between non-core
   gateways without degrading the degree of robustness relative to a
   mutually agreed security model. One way of doing this is is to
   extend the existing single-core autonomous-system model to include
   multiple core systems. This requires both a topological model which
   can be used to define the scope of these systems together with a
   global, trusted metric that can be used to drive the routing
   computations. An appropriate topological model is described in the
   next section, while an appropriate metric is suggested in the
   following section.

   2.1. Topological Models

      An "autonomous system" consists of a set of gateways, each of
      which can reach any other gateway in the same system using paths
      via gateways only in that system. The gateways of a system
      cooperatively maintain a routing data base using an interior
      gateway protocol (IGP) and a intra-system trusted routing
      mechanism of no further concern here. The IGP is expected to
      include security mechanisms to insure that only gateways of the
      same system can acquire each other as neighbors.

      One or more gateways in an autonomous system can run EGP with one
      or more gateways in a neighboring system. There is no restriction
      on the number or configuration of EGP neighbor paths, other than
      the requirement that each path involve only gateways of one system
      or the other and not intrude on a third system. It is
      specifically not required that EGP neighbors share a common
      network, although most probably will.

      An "autonomous confederation" consists of a set of autonomous
      systems sharing a common security model; that is, they trust each
      other to compute routes to other systems in the same
      confederation. Each gateway in a confederation can reach any
      other gateway in the same confederation using paths only in that
      confederation. Although there is no restriction on the number or
      configuration of EGP paths other than the above, it is expected
      that some mechanism be available so that potential EGP neighbors
      can discover whether they are in the same confederation. This
      could be done by access-control lists, for example, or by
      partitioning the set of system numbers.

      A network is "directly reachable" from an autonomous system if a
      gateway in that system has an interface to it. Every gateway in

that system is entitled to list all directly reachable networks in
EGP messages sent to any other system. In general, it may happen
that a particular network is directly reachable from more than one
system.

A network is "reachable" from an autonomous system if it is
directly reachable from an autonomous system belonging to the same
confederation. A directly reachable net is always reachable from
the same system. Every gateway in that confederation is entitled
to list all reachable nets in EGP messages sent to any other
system. It may happen that a particular net is either directly
reachable or reachable from different confederations.

In order to preserve global routing stability in the Internet, it
is explicitly assumed that routes within an autonomous system to a
directly reachable net are always preferred over routes outside
that system and that routes within an autonomous confederation are
always preferred over routes outside that confederation. The
mechanism by which this is assured is described in the next
section.

In general, EGP Update messages can include two lists of gateways,
one for those gateways belonging to the same system (internal
neighbors) and the other for gateways belonging to different
systems (external neighbors). Directly reachable nets must always
be associated with gateways of the same system, that is, with
internal neighbors, while non-directly reachable nets can be
associated with either internal or external neighbors. Nets that
are reachable, but not directly reachable, must always be
associated with gateways of the same confederation.

2.2. Trusted Routing Metrics

There seems to be a general principle which characterizes
distributed systems: The "nearer" a thing is the more dynamic and
trustable it is, while the "farther" a thing is the more static
and suspicious it is. For instance, the concept of network is
intrinsic to the Internet model, as is the concept of gateways
which bind them together. A cluster of gateways "near" each other
(e.g. within an autonomous system) typically exchange routing
information using a high-performance routing algorithm capable of
sensitive monitoring of, and rapid adaptation to, changing
performance indicators such as queueing delays and link loading.

However, clusters of gateways "far" from each other (e.g. widely
separated autonomous systems) usually need only coarse routing
information, possibly only "hints" on the best likely next hop to

the general destination area. On the other hand, mutual suspicion
increases with distance, so these clusters may need elaborate
security considerations, including peer authentication,
confidentiality, secrecy and signature verification. In addition,
considerations of efficiency usually dictate that the allowable
network bandidth consumed by the routing protocol itself decreases
with distance. The price paid for both of these things typically
is in responsiveness, with the effect that the more distant
clusters are from each other, the less dynamic is the routing
algorithm.

The above observations suggest a starting point for the evolution
of a globally acceptable routing metric. Assume the metric is
represented by an integer, with low values representing finer
distinctions "nearer" the gateway and high values coarser
distinctions "farther" from it. Values less than a globally
agreed constant X are associated with paths confined to the same
autonomous system as the sender, values greater than X but less
than another constant Y with paths confined to the autonomous
confederation of the sender and values greater than Y associated
with the remaining paths.

At each of these three levels - autonomous system, autonomous
confederation and universe of confederations - multiple routing
algorithms could be operated simultaneously, with each producing
for each destination net a possibly different subtree and metric
in the ranges specified above. However, within each system the
metric must have the same interpretation, so that other systems
can mitigate routes between multiple gateways in that system.
Likewise, within each confederation the metric must have the same
interpretation, so that other confederations can mitigate routes
to gateways in that confederation. Although all confederations
must agree on a common universe-of-confederations algorithm, not
all confederations need to use the same confederation-level
algorithm and not all systems in the same confederation need to
use the same system-level algorithm.

3. Implementation Issues

The manner in which the eight-bit "hop count" or distance field in
the EGP Update to be used is not specified in RFC-904, but left as a
matter for further study. The above model provides both an
interpretation of this field, as well as hints on how to design
appropriate routing algorithms.

For the sake of illustration, assume the values of X and Y above are
128 and 192 respectively. This means that the gateways in a

particular system will assign distance values less than 128 for
directly-reachable nets and that exterior gateways can compare these
values freely in order to select among these gateways. It also means
that the gateways in all systems of a particular confederation will
assign distance values between 128 and 192 for those nets not
directly reachable in the system but reachable in the confederation.
In the following it will be assumed that the various confederations
can be distinguished by some feature of the 16-bit system-number
field, perhaps by reserving a subfield.

3.1. Data-Base Management Functions

   The following implementation model may clarify the above issues,
   as well as present at least one way to organize the gateway data
   base. The data base is organized as a routing table, the entries
   of which include a net number together with a list of items, where
   each item consists of (a) the gateway address, system number and
   distance provided by an EGP neighbor, (b) a time-to-live counter,
   local routing information and other information as necessary to
   manage the data base.

   The routing table is updated each time an EGP Update message is
   received from a neighbor and possibly by other means, such as the
   system IGP. The message is first decoded into a list of quads
   consisting of a network number, gateway address, system number and
   distance. If the gateway address is internal to the neighbor
   system, as determined from the EGP message, the system number of
   the quad is set to that system; while, if not, the system number
   is set to zero, indicating "external."

   Next, a new value of distance is computed from the old value
   provided in the message and subject to the following constraints:
   If the system number matches the local system number, the new
   value is determined by the rules for the system IGP but must be
   less than 128. If not and either the system number belongs to the
   same confederation or the system number is zero and the old
   distance is less than 192, the value is determined by the rules
   for the confederation EGP, but must be at least 128 and less than
   192. Otherwise, the value is determined by the rules for the
   (global) universe-of-federations EGP, but must be at least 192.

   For each quad in the list the routing table is first searched for
   matching net number and a new entry made if not already there.
   Next, the list of items for that net number is searched for
   matching gateway address and system number and a new entry made if
   not already there. Finally, the distance field is recomputed, the
   time-to-live field reset and local routing information inserted.

The time-to-live fields of all items in each list are incremented
on a regular basis. If a field exceeds a preset maximum, the item
is discarded; while, if all items on a list are discarded, the
entire entry including net number is discarded.

When a gateway sends an EGP Update message to a neighbor, it must
invert the data base in order by gateway address, rather than net
number. As part of this process the routing table is scanned and
the gateway with minimum distance selected for each net number.
The resulting list is sorted by gateway address and partitioned on
the basis of internal/external system number.

3.2. Routing Functions

A gateway encountering a datagram (service unit) searches the
routing table for matching destination net number and then selects
the gateway on that list with minimum distance. As the result of
the value assignments above, it should be clear that routes at a
higher level will never be chosen if routes at a lower level
exist. It should also be clear that route selection within a
system cannot affect route selection outside that system, except
through the intervention of the intra-confederation routing
algorithm. If a simple min-system-hop algorithm is used for the
confederation EGP, the IGP of each system can influence it only to
the extent of reachability.

3.3. Compatibility Issues

The proposed interpretation is backwards-compatibile with known
EGP implementations which do not interpret the distance field and
with several known EGP implementations that take private liberties
with this field. Perhaps the simplest way to evolve the present
system is to collect the existing implementations that do not
interpet the distance field at all as a single confederation with
the present core system and routing restrictions. All distances
provided by this confederation would be assumed equal to 192,
which would provide at least a rudimentary capability for routing
within the universe of confederations.

One or more existing or proposed systems in which the distance
field has a uniform interpretation throughout the system can be
organized as autonomous confederations. This might include the
Butterfly gateways now now being deployed, as well as clones
elsewhere. These systems provide the capability to select routes
into the system based on the distance fields for the different
gateways. It is anticipated that the distance fields for the
Butterfly system can be set to at least 128 if the routing

Mills   [Page 9]

        information comes from another Butterfly system and to at least
        192 if from a non-Butterfly system presumed outside the
        confederation.

        New systems using an implmentation model such as suggested above
        can select routes into a confederation based on the distance
        field. For this to work properly, however, it is necessary that
        all systems and confederations adopt a consistent interpretation
        of distance values exceeding 192.

4.   Summary and Conclusions

     Taken at face value, this document represents a proposal for an
     interpretation of the distance field of the EGP Update message, which
     has previously been assigned no architected interpretation, but has
     been often used informally. The proposal amounts to ordering the
     autonomous systems in a hierarchy of systems and confederations,
     together with an interpretation of the distance field as a
     three-level metric. The result is to create a corresponding
     three-level routing community, one prefering routes inside a system,
     a second preferring routes inside a confederation and the third with
     no preference.

     While the proposed three-level hierarchy can readily be extended to
     any number of levels, this would create strain on the distance field,
     which is limited to eight bits in the current EGP model.

     The concept of distance can easily be generalized to "administrative
     distance" as suggested by John Nagle and others.

5.   References

     [1] Rosen, E., Exterior Gateway Protocol (EGP), DARPA Network
          Working Group Report RFC-827, Bolt Beranek and Newman, September
          1982.

     [2] Seamonson, L.J., and E.C., Rosen. "STUB" Exterior Gateway
          Protocol, DARPA Network Working Group Report RFC-888, BBN
          Communications, January 1984.

     [3] Mills, D.L., Exterior Gateway Protocol Formal Specification,
          DARPA Network Working Group Report RFC-904, M/A-COM Linkabit,
          April 1984.