          IPng White Paper on Transition and Other Considerations

Status of this Memo

Abstract

   This document was submitted to the IETF IPng area in response to RFC
   1550. Publication of this document does not imply acceptance by the
   IPng area of any ideas expressed within. Comments should be
   submitted to the big-internet@munnari.oz.au mailing list.

Summary

   This white paper outlines some general requirements for IPng in
   selected areas. It identifies the following requirements for stepwise
   transition:

   A) Interworking at every stage and every layer.
   B) Header translation considered harmful
   C) Coexistence.
   D) IPv4 to IPng address mapping.
   E) Dual stack hosts.
   F) DNS.
   G) Smart dual-stack code.
   H) Smart management tools.

   Some remarks about phsysical and logical multicast follow, and it is
   suggested that a model of how IPng will run over ATM is needed.

   Finally, the paper suggests that the requirements for policy routing,
   accounting, and security firewalls will in turn require all IPng
   packets to carry a trace of the type of transaction involved as well
   as of their source and destination.

Transition and deployment

   It is clear that the transition will take years and that every site
   will have to decide its own staged transition plan. Only the very
   smallest sites could envisage a single step ("flag day") transition,

Carpenter   [Page 1]

presumably under pressure from their Internet service providers.
Furthermore, once the IPng decision is taken, the next decade (or
more) of activity in the Internet and in all private networks using
the Internet suite will be strongly affected by the process of IPng
deployment. User sites will look at the decision whether to change
from IPv4 in the same way as they have looked in the past at changes
of programming language or operating system. It may not be a foregone
conclusion that what they change to is IPng. Their main concern will
be to minimise the cost of the change and the risk of lost
production.

This concern immediately defines strong constraints on the model for
transition and deployment of IPng. Some of these constraints are
listed below, with a short explanation of each one.

Terminology: an "IPv4 host" is a host that runs exactly what it runs
today, with no maintenance releases and no configuration changes. An
"IPng host" is a host that has a new version of IP, and has been
reconfigured. Similarly for routers.

A) Interworking at every stage and every layer.

This is the major constraint. Vendors of computer systems, routers,
and applications software will certainly not coordinate their product
release dates. Users will go on running their old equipment and
software. Therefore, any combination of IPv4 and IPng hosts and
routers must be able to interwork (i.e., participate in UDP and TCP
sessions). An IPv4 packet must be able to find its way from any IPv4
host, to any other IPv4 or IPng host, or vice versa, through a
mixture of IPv4 and IPng routers, with no (zero, null) modifications
to the IPv4 hosts. IPv4 routers must need no modifications to
interwork with IPng routers. Additionally, an application package
which is "aware" of IPv4 but still "unaware" of IPng must be able to
run on a computer system which is running IPv4, but communicating
with an IPng host. For example an old PC in Europe should be able to
access a NIC server in the USA, even if the NIC server is running
IPng and the transatlantic routing mechanisms are only partly
converted. Or a Class C network in one department of a company
should retain full access to corporate servers which are running
IPng, even though nothing whatever has been changed inside the Class
C network.

(This does NOT require an IPv4-only application to run on an IPng
host; thus we accept that some hosts cannot be upgraded until all
their applications are IPng-compatible. In other words we accept that
the API may change to some extent. However, even this relaxation is
debatable and some vendors may want to strictly preserve the IPv4 API
on an IPng host.)

B) Header translation considered harmful.

This author believes that any transition scenario which REQUIRES
dynamic header translation between IPv4 and IPng packets will create
almost insurmountable practical difficulties:

   B1) It is taken for granted for the purposes of this paper that
       IPng functionality will be a superset of IPv4 functionality.
       However, successful translation between protocols requires
       that the functionalities of the two protocols which are to be
       translated are effectively identical. To achieve this,
       applications will need to know when they are interworking,
       via the IPng API and a translator somewhere in the network,
       with an IPv4 host, so as to use only IPv4 functionality. This
       is an unrealistic constraint.

   B2) Administration of translators will be quite impracticable for
       large sites, unless the translation mechanism is completely
       blind and automatic. Specifically, any translation mechanism
       that requires special tags to be maintained manually for each
       host in tables (such as DNS tables or router tables) to
       indicate the need for translation will be impossible to
       administer. On a site with thousands of hosts running many
       versions and releases of several operating systems, hosts
       move forwards and even backwards between software releases in
       such a way that continuously tracking the required state of
       such tags will be impossible. Multiplied across the whole
       Internet, this will lead to chaos, complex failure modes, and
       difficult diagnosis. In particular, it will make the
       constraint of paragraph B1) impossible to respect.

       In practice, the knowledge that translation is needed should
       never leak out of the site concerned if chaos is to be
       avoided, and yet without such knowledge applications cannot
       limit themselves to IPv4 functionality when necessary.

To avoid confusion, note that header translation, as discussed here,
is not the same thing as address translation (NAT). This paper does
not discuss NAT.

This paper does not tackle performance issues in detail, but clearly
another disadvantage of translation is the consequent overhead.

C) Coexistence.

The Internet infrastructure (whether global or private) must allow
coexistence of IPv4 and IPng in the same routers and on the same

physical paths.

This is a necessity, in order that the network infrastructure can be
updated to IPng without requiring hosts to be updated in lock step
and without requiring translators.

Note that this requirement does NOT impose a decision about a common
or separate (ships-in-the-night) approach to routing. Nor does it
exclude encapsulation as a coexistence mechanism.

D) IPv4 to IPng address mapping.

Human beings will have to understand what is happening during
transition. Although auto-configuration of IPng addresses may be a
desirable end point, management of the transition will be greatly
simplified if there is an optional simple mapping, on a given site,
between IPv4 and IPng addresses.

Therefore, the IPng address space should include a mapping for IPv4
addresses, such that (if a site or service provider wants to do this)
the IPv4 address of a system can be transformed mechanically into its
IPng address, most likely by adding a prefix. The prefix does not
have to be the same for every site; it is likely to be at least
service-provider specific.

This does not imply that such address mapping will be used for
dynamic translation (although it could be) or to embed IPv4 routing
within IPng routing (although it could be). Its main purpose is to
simplify transition planning for network operators.

By the way, this requirement does not actually assume that IPv4
addresses are globally unique.

Neither does it help much in setting up the relationship, if any,
between IPv4 and IPng routing domains and hierarchies. There is no
reason to suppose these will be in 1:1 correspondence.

E) Dual stack hosts.

Stepwise transition without translation is hard to imagine unless a
large proportion of hosts are simultaneously capable of running IPng
and IPv4. If A needs to talk to B (an IPng host) and to C (an IPv4
host) then either A or B must be able to run both IPv4 and IPng. In
other words, all hosts running IPng must still be able to run IPv4.
IPng-only hosts are not allowed during transition.

This requirement does not imply that IPng hosts really have two
completely separate IP implementations (dual stacks and dual APIs),

but just that they behave as if they did. It is compatible with
encapsulation (i.e., one of the two stacks encapsulates packets for
the other).

Obviously, management of dual stack hosts will be simplified by the
address mapping just mentioned. Only the site prefix has to be
configured (manually or dynamically) in addition to the IPv4 address.

In a dual stack host the IPng API and the IPv4 API will be logically
distinguishable even if they are implemented as a single entity.
Applications will know from the API whether they are using IPng or
IPv4.

F) DNS.

The dual stack requirement implies that DNS has to reply with both an
IPv4 and IPng address for IPng hosts, or with a single reply that
encodes both.

If a host is attributed an IPng address in DNS, but is not actually
running IPng yet, it will appear as a black hole in IPng space - see
the next point.

G) Smart dual-stack code.

The dual-stack code may get two addresses back from DNS; which does
it use? During the many years of transition the Internet will
contain black holes. For example, somewhere on the way from IPng host
A to IPng host B there will sometimes (unpredictably) be IPv4-only
routers which discard IPng packets. Also, the state of the DNS does
not necessarily correspond to reality. A host for which DNS claims to
know an IPng address may in fact not be running IPng at a particular
moment; thus an IPng packet to that host will be discarded on
delivery. Knowing that a host has both IPv4 and IPng addresses gives
no information about black holes. A solution to this must be proposed
and it must not depend on manually maintained information. (If this
is not solved, the dual stack approach is no better than the packet
translation approach.)

H) Smart management tools.

A whole set of management tools is going to be needed during the
transition. Why is my IPng route different from my IPv4 route? If
there is translation, where does it happen? Where are the black
holes? (Cosmologists would like the same tool :-) Is that host REALLY
IPng-capable today?...

Multicasts high and low

   It is taken for granted that multicast applications must be supported
   by IPng. One obvious architectural rule is that no multicast packet
   should ever travel twice over the same wire, whether it is a LAN or
   WAN wire. Failure to observe this would mean that the maximum number
   of simultaneous multicast transactions would be halved.

   A negative feature of IPv4 on LANs is the cavalier use of physical
   broadcast packets by protcols such as ARP (and various non-IETF
   copycats). On large LANs this leads to a number of undesirable
   consequences (often caused by poor products or poor users, not by the
   protcol design itself). The obvious architectural rule is that
   physical broadcast should be replaced by unicast (or at worst,
   multicast) whenever possible.

ATM

   The networking industry is investing heavily in ATM. No IPng proposal
   will be plausible (in the sense of gaining management approval)
   unless it is "ATM compatible", i.e., there is a clear model of how it
   will run over an ATM network. Although a fully detailed document such
   as RFC 1577 is not needed immediately, it must be shown that the
   basic model works.

   Similar remarks could be made about X.25, Frame Relay, SMDS etc. but
   ATM is the case with the highest management hype ratio today.

Policy routing and accounting

   Unfortunately, this cannot be ignored, however much one would like
   to. Funding agencies want traffic to flow over the lines funded to
   carry it, and they want to know afterwards how much traffic there
   was. Accounting information can also be used for network planning
   and for back-charging.

   It is therefore necessary that IPng and its routing procedures allow
   traffic to be routed in a way that depends on its source and
   destination in detail. (As an example, traffic from the Physics
   department of MIT might be required to travel a different route to
   CERN than traffic from any other department.)

   A simple approach to this requirement is to insist that IPng must
   support provider-based addressing and routing.

   Accounting of traffic is required at the same level of detail (or
   more, for example how much of the traffic is ftp and how much is
   www?).

Both of these requirements will cost time or money and may impact
more than just the IP layer, but IPng should not duck them.

Security Considerations

Corporate network operators, and campus network operators who have
been hacked a few times, take this more seriously than many protocol
experts. Indeed many corporate network operators would see improved
security as a more compelling argument for transition to IPng than
anything else.

Since IPng will presumably be a datagram protocol, limiting what can
be done in terms of end-to-end security, IPng must allow more
effective firewalls in routers than IPv4. In particular efficient
traffic barring based on source and destination addresses and types
of transaction is needed.

It seems likely that the same features needed to allow policy routing
and detailed accounting would be needed for improved firewall
security. It is outside the scope of this document to discuss these
features in detail, but it seems unlikely that they are limited to
implementation details in the border routers. Packets will have to
carry some authenticated trace of the (source, destination,
transaction) triplet in order to check for unwanted traffic, to allow
policy-based source routing, and/or to allow detailed accounting.
Presumably any IPng will carry source and destination identifiers in
some format in every packet, but identifying the type of transaction,
or even the individual transaction, is an extra requirement.

Disclaimer and Acknowledgements

This is a personal view and does not necessarily represent that of my
employer.

CERN has been through three network transitions in recent years (IPv4
renumbering managed by John Gamble, AppleTalk Phase I to Phase II
transition managed by Mike Gerard, and DECnet Phase IV to DECnet/OSI
routing transition managed by Denise Heagerty). I could not have
written this document without having learnt from them. I have also
benefitted greatly from discussions with or the writings of many
people, especially various members of the IPng Directorate. Several
Directorate members gave comments that helped clarify this paper, as
did Bruce L Hutfless of Boeing. However the opinions are mine and
are not shared by all Directorate members.

Author's Address

    Brian E. Carpenter
    Group Leader, Communications Systems
    Computing and Networks Division
    CERN
    European Laboratory for Particle Physics
    1211 Geneva 23, Switzerland

    Phone: +41 22 767-4967
    Fax: +41 22 767-7155
    Telex: 419000 cer ch
    EMail: brian@dxcoms.cern.ch