NGTrans Working Group
INTERNET-DRAFT
Expires December 2002

Dave Thaler
Microsoft

29 June 2002

Support for Multicast over 6to4 Networks
<draft-ietf-ngtrans-6to4-multicast-01.txt>

Status of this Memo

This document is an Internet-Draft and is in full conformance with all provisions of Section 10 of RFC2026.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

Copyright Notice

1. Abstract

6to4 Tunneling allows isolated IPv6 domains or hosts, attached to an IPv4 network which has no native IPv6 support, to communicate with other such IPv6 domains or hosts with minimal manual configuration. This document defines support for IPv6 Multicast over 6to4 networks.

2. Introduction

6to4 Tunneling [6TO4] allows isolated IPv6 domains or hosts, attached to an IPv4 network which has no native IPv6 support, to communicate with other such IPv6 domains or hosts with minimal manual configuration. Effectively it treats the IPv4 network as a link layer. Since [6TO4] does not define support for multicast, the purpose of this document is to define such support.

Unlike [6OVER4], 6to4 does not assume the general availability of wide-area IPv4 multicast. The 6to4 mechanism assumes only unicast capability in its underlying IPv4 carrier network. As a result, the IPv4 network appears as a large Non-Broadcast Multi-Access (NBMA) link, over which we require the ability to multicast. To do this, IPv6 multicast packets being sent to or from a 6to4 router must be encapsulated in IPv4 unicast packets. If the tree has multiple branches in the 6to4 address space, 6to4 encapsulation of the same multicast packet will take place multiple times by necessity.

In this document, we refer to the device in a 6to4 site which has a 6to4 pseudo-interface as a 6to4 "gateway". The gateway may either be a host (if the site is just a single host), or a router (if the site includes IPv6 links). We use the term "relay" as defined in [6TO4].

3. Overview

We assume each 6to4 gateway points an IPv6 default route at a particular relay reachable across the IPv4 infrastructure. Each relay, plus the set of all gateways (perhaps unknown to the relay) using the relay, together can be thought of as being on a separate

NBMA link.

All IPv6 multicast packets will be encapsulated in IPv4 unicast packets over the logical NBMA link. This requires the 6to4 relay to keep state for each gateway which has joined a particular group. IPv6 multicast packets from the IPv6 native infrastructure behind the relay will be sent to each gateway which has requested them.

Since the number of gateways using a relay can be quite large, and we expect that most sites will not want to receive most groups, an explicit-joining protocol is required for gateways to communicate group membership information to a relay. The two most likely candidates are the Multicast Listener Discovery [MLD] protocol, and the PIM-Sparse Mode [PIMSM] protocol. Since a 6to4 gateway may be a host, and hosts typically do not implement routing protocols, gateways will use MLD as described in Section 4 below. This allows a host kernel to easily implement 6to4 gateway behavior, and obviates the relay from the need to know whether a given gateway is a host or a router. From the relay's perspective, all gateways are indistinguishable from hosts on an NBMA leaf network.

All IPv6 multicast packets sourced within the 6to4 site (and sent to a scope larger than the 6to4 site) are forwarded by the 6to4 gateway to the relay over the 6to4 pseudo-interface, regardless of whether any remote receivers exist. This is done so that the gateway (which may be a host) need not be aware of external membership information.

When a relay receives a multicast packet via 6to4 encapsulation, it applies a Reverse-Path Forwarding check by dropping it unless the source address is a 6to4 address. If it is not dropped, the packet is forwarded to all other 6to4 gateways which have requested it, in addition to being forwarded out any native IPv6 interfaces according to the rules of the multicast routing protocol(s) running on the relay. When applying the rules for multicast interoperability [INTEROP], the 6to4 link is treated using the MLD equivalents of the rules for an IGMP-only link.

4. Multicast Listener Discovery

The MLD protocol usually operates by having the Querier multicast an MLD Query message on the link. This behavior does not work on NBMA links which do not support multicast. Since the set of gateways is typically unknown to the relay (and potentially quite

large), unicasting the queries is also impractical. The following behavior is used instead.

The relay operates passively, sending no Queries but simply tracking membership information according to Reports and Done messages. To provide robustness, gateways unicast Reports to the relay every [Query Interval] (defined as 125 in [MLD]) seconds. The IPv6 source address of MLD reports sent to a 6to4 relay MUST be a link-local address formed as specified in Section 3.7 of [MECH].

Reports are not relayed to other gateways by the 6to4 relay, and no report suppression is done. As a result, the timer used to send periodic reports MUST be initialized to a random value from the interval [0, [Query Interval]] before sending the first periodic report, in order to prevent startup synchronization (e.g., after a power outage).

If a gateway is serving as a local router, it SHOULD also function as an "MLD Proxy". MLD Proxy behavior can be summarized as follows. First, the gateway will serve as an MLD querier on its private interface(s), and send MLD reports to the relay for groups which are scoped larger than a site and have members present on its private interface(s). Second, the gateway will forward multicast packets received encapsulated from a relay to any private interface with members present.

5. Scalability Considerations

The requirement that a relay keep group state per gateway that has

joined the group introduces potential scalability concerns. In this section, we discuss how these concerns may be addressed.

Scalability of 6to4 can be achieved by adding more relays, and using an appropriate relay discovery mechanism for gateways to discover relays. Since there are non-multicast related reasons, such as bandwidth bottlenecks at relays, to add more relays as well, a detailed discussion of relay discovery is outside the scope of this document.

There is, however, work in progress on this topic. One solution is be to use a well-known DNS name, and have gateways periodically (e.g. once a day) re-resolve the DNS name and update their relay's address to use accordingly. A potentially better solution is to

assign an IPv4 anycast address to relays (e.g., [ANYCAST]). However, sending periodic MLD Reports to an anycast address can cause duplicates. Specifically, if routing changes such that a different relay receives a periodic MLD Report, both the new and old relays will encapsulate data to the 6to4 site until the old relay's state times out. This is obviously undesirable.

Hence, if a gateway is aware that a relay's IPv4 address is an anycast address (such as because it is well-known), then the gateway MUST determin an IPv4 unicast address of a relay and use that instead for all MLD encapsulation. The recommended procedure is that a gateway having an anycast address of a relay should send an ICMPv4 Echo Request to the anycast address, and that relays should respond with an Echo Reply from their unicast address (since anycast addresses should not be used as source addresses). These "pings" may be done periodically (e.g. once a day) to re-resolve the unicast address of a close relay.

Since adding another relay has the result of adding another independent NBMA link, this allows the gateways to be spread out among more relays so as to keep the number of relays per gateway at a reasonable level.

6. Supporting Multiple Relays in the Presence of RPF

A problem with the simple mechanisms described above can occur when multiple relays exist, and a multicast routing protocol is used that employs Reverse-Path Forwarding (RPF) checks against the source address, such as occurs when [PIMSM] is used by the IPv6 infrastructure. Namely, if an IPv6 router on the path to a receiver in the native IPv6 infrastructure expects to receive data from a 6to4 site via the closest relay to the receiver, that relay may not be the one to which the 6to4 site is encapsulating data, and no data will be seen.

The solution specified by this document is that if a relay receives an explicit join from the native IPv6 infrastructure, for a given source and group pair where the source is a 6to4 address, then the relay will periodically (using the same rules in Section 4) unicast an MLD report for the group to the 6to4 site gateway. The 6to4 gateway must keep state per relay from which an MLD Report has been sent, and in addition to forwarding multicast traffic from the site to its relay of choice, traffic is also

forwarded to all relays from which Reports have been received. The choice of whether this state and replication is done at at the link-layer (i.e., by the tunnel interface) or at the network-layer is implementation-dependent.

The solution above will scale to an arbitrary number of relays, as long at the number of relays requiring multicast traffic from a given 6to4 site remains reasonable enough to not overly burden the site gateway. (Note that bi-directional tree protocols such as BGMP [BGMP] do not use RPF checks, and so would prevent the problem from occurring to begin with.)

7. Supporting Site-Site Multicast

Since we require gateways to accept MLD unicast reports, as described above, it is also possible to support multicast among 6to4 sites, without requiring assistance from any relays, in two cases.

First, when a site gateway wants to join a given (source, group) pair, where the source is another 6to4 address, then it periodically unicast encapsulates an MLD Report to the site gateway for the source.

The second case is when it wants to join a unicast-prefix-based multicast addresses [UNIMCAST], and the unicast prefix embedded in the multicast address is another site's 6to4 prefix. In this case, it periodically unicast encapsulates an MLD Report to the site gateway for that prefix.

For all other types of joins, the gateway periodically sends MLD Reports to the relay, as discussed in section 4.

We note that this can result in a significant amount of state at a site gateway sourcing multicast to a large number of other 6to4 sites. However, it is expected that this is not unreasonable for two reasons. First, the 6to4 most likely does not have native multicast connectivity (or else it could use 6over4 instead), and as a result is likely doing unicast replication at present. The amount of state is thus the same as what such a site already deals with. Secondly, any site expecting to source traffic to a large number of sites could get a point-to-point tunnel to the native IPv6 infrastruction, and use that instead of 6to4.

8. Security Considerations

The same security considerations and solutions discussed in [6TO4] apply to multicast traffic.

9. Acknowledgements

The following individuals provided helpful discussion on the mechanisms in this document: Rich Draves, Brian Zill, Steve Deering, and Brian Carpenter.

10. Author's Address

Dave Thaler
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052-6399
Phone: +1 425 703 8835
EMail: dthaler@microsoft.com

11. Normative References

[6TO4]
Carpenter, B., and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[INTEROP]
D. Thaler, "Interoperability Rules for Multicast Routing Protocols", RFC 2715, October 1999.

[MECH]
Gilligan, R., and E. Nordmark, "Transition Mechanisms for IPv6 Hosts and Routers", RFC 2893, August 2000.

[MLD]
Deering, S., Fenner, W., and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, October 1999.

[UNIMCAST]
Haberman, B., and D. Thaler, "Unicast-Prefix-based IPv6 Multicast Addresses", Work in progress, draft-ietf-ipngwg-uni-based-mcast-03.txt, October 2001.

12. Non-Normative References

[6OVER4]
Carpenter, B., and C. Jung, "Transmission of IPv6 over IPv4 Domains without Explicit Tunnels", RFC 2529, March 1999.

[ANYCAST]

C. Huitema, "An anycast prefix for 6to4 relay routers", RFC 3068, June 2001.

[BGMP]
D. Thaler, "Border Gateway Multicast Protocol (BGMP): Protocol Specification", draft-ietf-bgmp-spec-03.txt, Work in progress, June 2002.

[PIMSM]
Estrin, D. Farinacci, D., Helmy, A., Thaler, D., Deering, S., Handley, M., Jacobson, V., Liu, C., Sharma, P., and L. Wei. "Protocol Independent Multicast-Sparse Mode (PIM-SM): Protocol Specification", RFC 2362, June 1998.

13. Full Copyright Statement