Network Working Group
Request for Comments: 2491
Category: Standards Track

G. Armitage
Lucent Technologies
P. Schulter
Bright Tiger Technologies
M. Jork
Digital Equipment GmbH
G. Harter
Compaq

January 1999

IPv6 over Non-Broadcast Multiple Access (NBMA) networks

Status of this Memo

Copyright Notice

Abstract

This document describes a general architecture for IPv6 over NBMA networks. It forms the basis for subsidiary companion documents that describe details for various specific NBMA technologies (such as ATM or Frame Relay). The IPv6 over NBMA architecture allows conventional host-side operation of the IPv6 Neighbor Discovery protocol, while also supporting the establishment of 'shortcut' NBMA forwarding paths when dynamically signaled NBMA links are available. Operations over administratively configured Point to Point NBMA links are also described.

Dynamic NBMA shortcuts are achieved through the use of IPv6 Neighbor Discovery protocol operation within Logical Links, and inter-router NHRP for the discovery of off-Link NBMA destinations. Both flow-triggered and explicitly source-triggered shortcuts are supported.

1. Introduction.

Non Broadcast Multiple Access (NBMA) networks may be utilized in a variety of ways. At one extreme, they can be used to simply provide administratively configurable point to point service, sufficient to interconnect IPv6 routers (and even IPv6 hosts, in certain

situations). At the other extreme, NBMA networks that support dynamic establishment and teardown of Virtual Circuits (or functional equivalents) may be used to emulate the service provided to the IPv6 layer by conventional broadcast media such as Ethernet. Typically this emulation requires complex convergence protocols, particularly to support IPv6 multicast.

This document describes a general architecture for IPv6 over NBMA networks. It forms the basis for companion documents that provide details specific to various NBMA technologies (for example, ATM [17] or Frame Relay). The IPv6 over NBMA architecture allows conventional host-side operation of the IPv6 Neighbor Discovery protocol, while also supporting the establishment of 'shortcut' NBMA forwarding paths (when dynamically signaled NBMA links are available).

The majority of this document focuses on the use of dynamically managed point to point and point to multipoint calls between interfaces on an NBMA network. These will be generically referred to as "SVCs" in the rest of the document. The use of administratively configured point to point calls will also be discussed. Such calls will be generically referred to as "PVCs". Depending on context, either may be shortened to "VC".

Certain NBMA networks may provide a form of connectionless service (e.g. SMDS). In these cases, a "call" or "VC" shall be considered to implicitly exist if the sender has an NBMA destination address to which it can transmit packets whenever it desires.

1.1 Neighbor Discovery.

A key difference between this architecture and previous IP over NBMA protocols is its mechanism for supporting IPv6 Neighbor Discovery.

The IPv4 world evolved an approach to address resolution that depended on the operation of an auxiliary protocol operating at the 'link layer' - starting with Ethernet ARP (RFC 826 [14]). In the world of NBMA (Non Broadcast, Multiple Access) networks ARP has been applied to IPv4 over SMDS (RFC 1209 [13]) and IPv4 over ATM (RFC 1577 [3]). More recently the ION working group has developed NHRP (Next Hop Resolution Protocol [8]), a general protocol for performing intra-subnet and inter-subnet address resolution applicable to a range of NBMA network technologies.

IPv6 developers opted to migrate away from a link layer specific approach, chosing to combine a number of tasks into a protocol known as Neighbor Discovery [7], intended to be non-specific across a number of link layer technologies. A key assumption made by Neighbor Discovery's actual protocol is that the link technology underlying a

given IP interface is capable of native multicasting. This is not particularly true of most NBMA network services, and usually requires convergence protocols to emulate the desired service. (The MARS protocol, RFC 2022 [5], is an example of such a convergence protocol.) This document augments and optimizes the MARS protocol for use in support of IPv6 Neighbor Discovery, generalizing the applicability of RFC 2022 beyond ATM networks.

1.2 NBMA Shortcuts.

A shortcut is an NBMA level call (VC) directly connecting two IP endpoints that are logically separated by one or more routers at the IP level. IPv6 packets traversing this VC are said to 'shortcut' the routers that are in the logical IPv6 path between the VC's endpoints.

NBMA shortcuts are a mechanism for minimizing the consumption of resources within an IP over NBMA cloud (e.g. router hops and NBMA VCs).

It is important that NBMA shortcuts are supported whenever IP is deployed across NBMA networks capable of supporting dynamic establishment of calls (SVCs or functional equivalent). For IPv6 over NBMA, shortcut discovery and management is achieved through a mixture of Neighbor Discovery and NHRP.

1.3 Key components of the IPv6 over NBMA architecture.

1.3.1 NBMA networks providing PVC support.

When the NBMA network is used in PVC mode, each PVC will connect exactly two nodes and the use of Neighbor Discovery and other IPv6 features is limited. IPv6/NBMA interfaces have only one neighbor on each Link. The MARS and NHRP protocols are NOT necessary, since multicast and broadcast operations collapse down to an NBMA level unicast operation. Dynamically discovered shortcuts are not supported.

The actual details of encapsulations and link token generation SHALL be covered by companion documents covering specific NBMA technology. They SHALL conform to the following guidelines:

Both unicast and multicast IPv6 packets SHALL be transmitted over PVC links using the encapsulation described in section 4.4.1.

Interface tokens for PVC links SHALL be constructed as described in section 5. Interface tokens need only be unique between the two nodes on the PVC link.

This use of PVC links does not mandate, nor does it prohibit the use of extensions to the Neighbor Discovery protocol which may be developed for either general use of for use in PVC connections (for example, Inverse Neighbor Discovery).

NBMA-specific companion documents MAY additionally specify the concatenation of IPv6 over PPP and PPP over NBMA mechanisms as an OPTIONAL approach to point to point IPv6.

Except where noted above, the remainder of this document focuses on the SVC case.

1.3.2 NBMA networks providing SVC support.

When the NBMA network is used in SVC mode, the key components are:

- The IPv6 Neighbor model, where neighbors are discovered through the use of messages multicast to members of an IPv6 interface's local IPv6 Link.
- The MARS model, allowing emulation of general multicast using multipoint calls provided by the underlying NBMA network.
- The NHRP service for seeking out the NBMA identities of IP interfaces who are logically distant in an IP topological sense.
- The modeling of IP traffic as 'flows', and optionally using the existence of a flow as the basis for attempting to set up a shortcut link level connection.

In summary:

The IPv6 "Link" is generalized to "Logical Link" (LL) in NBMA environments (analogous to the generalization of IPv4 IP Subnet to Logical IP Subnet in RFC 1209 and subsequently RFC 1577).

IPv6/NBMA interfaces utilize RFC 2022 (MARS) for general intra-Logical Link multicasting. The MARS itself is used to optimally distribute discovery messages within the Logical Link.

For destinations not currently considered to be Neighbors, a host sends the packets to one of its default routers.

When appropriately configured, the egress router from a Logical Link is responsible for detecting the existence of an IP packet flow through it that might benefit from a shortcut connection.

While continuing to conventionally forward the flow's packets, the router initiates an NHRP query for the flow's destination IP address.

The last router/NHS before the target of the NHRP query ascertains the target interface's preferred NBMA address.

The originally querying router then issues a Redirect to the IP source, identifying the flow's destination as a transient Neighbor.

Host-initiated triggering of shortcut discovery, regardless of the existence of a packet flow, is also supported through specific Neighbor Solicitations sent to a source host's default router.

A number of key advantages are claimed for this approach. These are:

The IPv6 stacks on hosts do not implement separate ND protocols for each link layer technology.

When the destination of a flow is solicited as a transient neighbor, the returned NBMA address will be the one chosen by the destination when the flow was originally established through hop-by-hop processing. This supports the existing ND ability for IPv6 destinations to perform their own dynamic interface load sharing.

1.4 Terminology.

The bit-pattern or numeric value used to identify a particular NBMA interface at the NBMA level will be referred to as an "NBMA address". (An example would be an ATM End System Address, AESA, when applying this architecture to ATM networks, or an E.164 number when applying this architecture to SMDS networks.)

The call that, once established, is used to transfer IP packets from one NBMA interface to another will be referred to as an SVC or PVC depending on whether the call is dynamically established through some signaling mechanism, or administratively established. The specific signaling mechanisms used to establish or tear down an SVC will be defined in the NBMA-specific companion specifications. Certain NBMA networks may provide a form of connectionless service (e.g. SMDS). In these cases, a "call" or "SVC" shall be considered to implicitly exist if the sender has an NBMA destination address to which it can transmit packets whenever it desires.

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [16].

1.5 Document Structure.

The remainder of this document is structured as follows: Section 2 explains the generalization of IPv6 Link to "Logical Link" when used over NBMA networks, and introduces the notion of the Transient Neighbor. Section 3 describes the modifications to the MARS protocol for efficient distribution of ND messages within a Logical Link, and the rules and mechanisms for discovering Transient Neighbors. Section 4 covers the basic rules governing IPv6/NBMA interface initialization, packet and control message encapsulations, and rules for SVC management. Section 5 describes the general rules for constructing Interface Tokens, the Link Layer Address Option, and Link Local addresses. Section 6 concludes the normative sections of the document. Appendix A provides some non-normative descriptive text regarding the operation of Ipv6 Neighbor Discovery. Appendix B describes some sub-optimal solutions for emulating the multicasting of Neighbor Discovery messages around a Logical Link. Appendix C discusses shortcut suppression and briefly reviews the future relationships between flow detection and mapping of flows onto SVCs of differing qualities of service.

2. Logical Links, and Transient Neighbors.

IPv6 contains a concept of on-link and off-link. Neighbors are those nodes that are considered on-link and whose link-layer addresses may therefore be located using Neighbor Discovery. Borrowing from the terminology definitions in the ND text:

on-link - an address that is assigned to a neighbor’s interface on a shared link. A host considers an address to be on-link if:
  - it is covered by one of the link’s prefixes, or
  - a neighboring router specifies the address as the target of a Redirect message, or
  - a Neighbor Advertisement message is received for the target address, or
  - a Neighbor Discovery message is received from the address.

off-link - the opposite of "on-link"; an address that is not assigned to any interfaces attached to a shared link.

Off-link nodes are considered to only be accessible through one of the routers directly attached to the link.

The NBMA environment complicates the sense of the word 'link' in much the same way as it complicated the sense of 'subnet' in the IPv4 case. For IPv4 this required the definition of the Logical IP Subnet (LIS) - an administratively constructed set of hosts that would share the same routing prefixes (network and subnetwork masks).

This document considers the IPv6 analog to be a Logical Link (LL).

An LL consists of nodes administratively configured to be 'on link' with respect to each other.

The members of an LL are an IPv6 interface's initial set of neighbors, and each interface's Link Local address only needs to be unique amongst this set.

It should be noted that whilst members of an LL are IPv6 Neighbors, it is possible for Neighbors to exist that are not, administratively, members of the same LL.

Neighbor Discovery events can result in the expansion of an IPv6 interface's set of Neighbors. However, this does not change the set of interfaces that make up its LL. This leads to three possible relationships between any two IPv6 interfaces:

- On LL, Neighbor.
- Off LL, Neighbor.
- Off LL, not Neighbor.

Off LL Neighbors represent the 'shortcut' connections, where it has been ascertained that direct connectivity at the NBMA level is possible to a target that is not a member of the source's LL.

Neighbors discovered through the operation of unsolicited messages, such as Redirects, are termed 'Transient Neighbors'.

3. Intra-LL and Inter-LL Discovery.

This document makes a distinction between the discovery of neighbors within a Logical Link (intra-LL) and neighbors beyond the LL (inter-LL). The goal is to allow both inter- and intra-LL neighbor discovery to involve no changes to the host-side IPv6 stack for NBMA interfaces.

Note that section 1.3.1 applies when the NBMA network is being used to provide only configured point to point (PVC) service.

3.1 Intra-LL - ND over emulated multicast.

The basic model of ND assumes that a link layer interface will do something meaningful with an ICMPv6 packet sent to a multicast IP destination address. (IPv6 assumes that multicasting is an integral part of the Internet service.) This document assumes multicast support will be provided using the RFC 2022 (MARS) [5] service (generalized for use over other NBMA technologies in addition to ATM). An IPv6 LL maps directly onto an IPv6 MARS Cluster in the same way an IPv4 LIS maps directly onto an IPv4 MARS Cluster.

The goal of intra-LL operation is that the IPv6 layer must be able to simply pass multicast ICMPv6 packets down to the IPv6/NBMA driver without any special, NBMA specific processing. The underlying mechanism for distributing Neighbor Discovery and Router Discovery messages then works as expected.

Sections 3.1.1 describes the additional functionality that SHALL be required of any MARS used in conformance with this document. Background discussion of these additions is provided in Appendix B.

3.1.1 Mandatory augmented MARS and MARS Client behavior.

IPv6/NBMA interfaces SHALL register as MARS Cluster members as described in section 4.1, and SHALL send certain classes of outgoing IPv6 packets directly to their local MARS as described in section 4.4.2.

The MARS itself SHALL then re-transmit these packets according to the following rules:

- When the MARS receives an IPv6 packet, it scans the group membership database to find the NBMA addresses of the IPv6 destination group’s members.

- The MARS then checks to see if every group member currently has its pt-pt control VC open to the MARS. If so, the MARS sends a copy of the data packet directly to each group member over the existing pt-pt VCs.

- If one or more of the discovered group members do not have an open pt-pt VC to the MARS, or if there are no group members listed, the packet is sent out ClusterControlVC instead. No copies of the packet are sent over the existing (if any) pt-pt VCs.

3.2 Inter-LL - Redirects, and their generation.

Shortcut connections are justified on the grounds that demanding flows of IP packets may exist between source/destination pairs that are separated by IP routing boundaries. Shortcuts are created between Transient Neighbors.

The key to creating transient neighbors is the Redirect message (section 8 [7]). IPv6 allows a router to inform the members of an LL that there is a better 'first hop' to a given destination (section 8.2 [7]). The advertisement itself is achieved through a Router Redirect message, which may carry the link layer address of this better hop.

A transmitting host only listens to Router Redirects from the router that is currently acting as the default router for the IP destination that the Redirect refers to. If a Redirect arrives that indicates a better first hop for a given destination, and supplies a link layer (NBMA) address to use as the better first hop, the associated Neighbor Cache entry in the source host is updated and its reachability set to STALE. Updating the cache in this context involves building a new VC to the new NBMA address. If this is successful, the old VC is torn down only if it no longer required (since the old VC was to the router, it may still be required by other packets from the host that are heading to the router).

Two mechanisms are provided for triggering the discovery of a better first hop:

Router-based flow identification/detection.

Host-initiated shortcut request.

Section 3.2.1 discusses flow-based triggers, section 3.2.2 discusses the host initiated trigger, and section 3.2.3 discusses the use of NHRP to discover mappings for IPv6 targets in remote LLs.

3.2.1 Flow Triggered Redirection.

The modification of forwarding paths based on the dynamic detection of IP packet flows is at the core of models such as the Cell Switch Router [11] and the IP Switch [12]. Responsibility for detecting flows is placed into the routers, where packets cross the edges of IP routing boundaries.

For the purpose of conformance with this document, a router MAY choose to initiate the discovery of a better first-hop when it determines that an identifiable flow of IP packets are passing through it.

Such a router:

SHALL only track flows that originate from a directly attached host (a host that is within the LL-local scope of one of the router's interfaces).

SHALL NOT use IP packets arriving from another router to trigger the generation of a Router Redirect.

SHALL only consider IPv6 packets with FlowID of zero for the purposes of flow detection as defined in this section.

SHALL utilize NHRP as described in section 3.2.3 to ascertain a better first-hop when a suitable flow is detected, and advertise the information in a Router Redirect.

IPv6 routers that support the OPTIONAL flow detection behavior described above SHALL support administrative mechanisms to switch off flow-detection. They MAY provide mechanisms for adding additional constraints to the categories of IPv6 packets that constitute a 'flow'.

The actual algorithm(s) for determining what sequence of IPv6 packets constitute a 'flow' are outside the scope of this document. Appendix C discusses the rationale behind the use of non-zero FlowID to suppress flow detection.

3.2.2 Host Triggered Redirection

A source host MAY also trigger a redirection to a transient neighbor. To support host-triggered redirects, routers conforming to this document SHALL recognize specific Neighbor Solicitation messages sent by hosts as requests for the resolution of off-link addresses.

To perform a host-triggered redirect, a source host SHALL:

Create a Neighbor Solicitation message referring to the off-LL destination (target) for which a shortcut is desired

Address the NS message to the router that would be the next hop for traffic sent towards the off-LL target (rather than the target's solicited node multicast address).

Use the standard ND hop limit of 255 to ensure the NS won't be discarded by the router.

Include the shortcut limit option defined in appendix D. The value of this option should be equal to the hop limit of the data flow for which this trigger is being sent. This ensures that the router is able to restrict the shortcut attempt to not exceed the reach of the data flow.

Forward the NS packet to the router that would be the next hop for traffic sent towards the off-LL target.

Routers SHALL consider a unicast NS with shortcut limit option as a request for a host-triggered redirect. However, actual shortcut discovery is OPTIONAL for IPv6 routers.

When shortcut discovery is not supported, the router SHALL construct a Redirect message identifying the router itself as the best 'shortcut', and return it to the soliciting host.

If shortcut discovery is to be supported, the router's response SHALL be:

A suitable NHRP Request is constructed and sent as described in section 3.2.3. The original NS message SHOULD be discarded.

Once the NHRP Reply is received by the originating router, the router SHALL construct a Redirect message containing the IPv6 address of the transient neighbor, and the NBMA link layer address returned by the NHRP resolution process.

The resulting Redirect message SHALL then be transmitted back to the source host. When the Redirect message is received, the source host SHALL update its Neighbor and Destination caches.

The off-LL target is now considered a Transient Neighbor. The next packet sent to the Transient Neighbor will result in the creation of the direct, shortcut VC (to the off-LL target itself, or to the best egress router towards that neighbor as determined by NHRP).

If a NHRP NAK or error indication is received for a host-triggered shortcut attempt, the requesting router SHALL construct a Redirect message identifying the router itself as the best 'shortcut', and return it to the soliciting host.

3.2.3 Use of NHRP between routers.

Once flow detection has occurred, or a host trigger has been detected, routers SHALL use NHRP in an NHS to NHS mode to establish the IPv6 to link level address mapping of a better first hop.

IPv6/NBMA routers supporting shortcut discovery will need to perform some or all of the following functions:

- Construct NHRP Requests and Replies.

- Parse incoming NHRP Requests and Replies from other NHSes (routers).

  - Forward NHRP Requests towards an NHS that is topologically closer to the IPv6 target.

  - Forward NHRP Replies towards an NHS that is topologically closer to the requester.

  - Perform syntax translation between Neighbor Solicitations and outbound NHRP Requests.

  - Perform syntax translation between inbound NHRP Replies and Redirects.

The destination of the flow that caused the trigger (or the target of the host initiated trigger) is used as the target for resolution in a NHRP Request. The router then forwards this NHRP Request to the next closest NHS. The process continues (as it would for normal NHRP) until the Request reaches an NHS that believes the IP target is within link-local scope of one of its interfaces. (This may potentially occur within a single router.)

As NHRP resolution requests always follow the routed path for a given target protocol address, the scope of a shortcut request will be automatically bounded to the scope of the IPv6 target address. (e.g. resolution requests for site-local addresses will not be forwarded across site boundaries.)

The last hop router SHALL resolve the NHRP Request from mapping information contained in its neighbor cache for the interface on which the specified target is reachable. If there is no appropriate entry in the Neighbor cache, or the destination is currently considered unreachable, the last hop router SHALL perform Neighbor Discovery on the local interface, and build the NHRP Reply from the resulting answer. (Note, in the case where the NHRP Request originated due to flow detection, there must already be a hop-by-hop

flow of packets going through the last hop router towards the target. In this typical case the Neighbor cache will already have the desired information.)

The NHRP Reply is propagated back to the source of the NHRP Request, using a hop-by-hop path as it would for normal NHRP.

If the discovery process was triggered through flow detection at the originating router, the return of the NHRP Reply results in the following events:

A Redirect is constructed using the IPv6/NBMA mapping carried in the NHRP Reply.

The Redirect is unicast to the IP packet flow's source (using the VC on which the flow is arriving at the router, if it is a bi-directional pt-pt VC).

Any Redirect message sent by a router MUST conform to all the rules described in [7] so that the packet is properly validated by the receiving host. Specifically, if the target of the resulting short-cut is the destination host then the ICMP Target Address MUST be the same as the ICMP Destination Address in the original message. If the target of the short-cut is an egress router then the ICMP Target Address MUST be a Link Local address of the egress router that is unique to the NBMA cloud to which the router's NBMA interface is attached.

Also note that egress routers may subsequently redirect the source host. To do so, the Link Local ICMP Source Address of the Redirect message MUST be the same as the Link Local ICMP Target Address of the original Redirect message.

Note that the router constructing the NHRP Reply does so using the NBMA address returned by the target host when the target host first accepted the flow of IP traffic. This retains a useful feature of Neighbor Discovery - destination interface load sharing.

Upon receipt of a NHRP NAK reply or error indication for a flow-triggered shortcut attempt, no indication is sent to the source of the flow.

3.2.3.1 NHRP/ND packet translation rules.

The following translation rules are meant to augment the packet format specification in section 5 of the NHRP specification [8], covering those packet fields specifically utilized by the IPv6/NBMA architecture.

NHRP messages are constructed and sent according to the rules in [8]. The value of the NBMA technology specific fields such as arafn, arpro.type, ar$pro.snap and link layer address format are defined in NBMA-specific companion documents. Source, destination or client protocol addresses in the common header or CIE of a NHRP message are always IPv6 addresses of length 16.

When constructing an host-triggered NHRP resolution request in response to a Neighbor Solicitation:

The ar$hopcnt field MUST be smaller than the shortcut limit value specified in the shortcut limit option included in the triggering NS message. This ensures that hosts have control over the reach of their shortcut request. Note that the shortcut limit given in the option is relative to the requesting host, thus the requirement of ar$hopcnt being smaller than the given shortcut limit.

The Flags field in the common header of the NHRP resolution request SHOULD have the Q and S bits set.

The U bit SHOULD be set. NBMA and protocol source addresses are those of the router constructing the request.

The target address from the NS message is used as the NHRP destination protocol address. A CIE SHALL NOT be specified.

When constructing a NHRP resolution request as a result of flow detection, the choice of values is configuration dependent.

A NHRP resolution reply is build according to the rules in [8].

For each CIE returned, the holding time is 10 minutes.

The MTU may be 0 or a value specified in the NBMA-specific companion document.

A successful NHRP resolution reply for a host-triggered shortcut attempt is translated into an IPv6 Redirect message as follows:

IP Fields:

Source Address
: The link-local address assigned to the router’s interface from which this message is sent.

Destination Address
: IPv6 Source Address of the triggering NS

Hop Limit
: 255

ICMP Fields:
Target Address
NHRP Client Protocol Address
Destination Address
Target of triggering NS (this is equivalent to the NHRP Destination Protocol Address)
Target link-layer address
NHRP Client NBMA Address

All NHRP extensions currently defined in [8] have no effect on NHRP/ND translation and MAY be used in NHRP messages for IPv6.

3.2.3.2 NHRP Purge rules.

Purges are generated by NHRP when changes are detected that invalidate a previously issued NHRP Reply (this may include topology changes, or a target host going down or changing identity). Any IPv6 shortcut previously established on the basis of newly purged information SHOULD be torn down.

Routers SHALL keep track of NHRP cache entries for which they have issued Neighbor Advertisements or Router Redirects. If a NHRP Purge is received that invalidates information previously issued to local host, the router SHALL issue a Router Redirect specifying the router itself as the new best next-hop for the affected IPv6 target.

Routers SHALL keep track of Neighbor cache entries that have previously been used to generate an NHRP Reply. The expiry of any such Neighbor cache entry SHALL result in a NHRP Purge being sent towards the router that originally requested the NHRP Reply.

3.3. Neighbor Unreachability Detection.

Neighbor Solicitations sent for the purposes of Neighbor Unreachability Detection (NUD) are unicast to the Neighbor in question, using the VC that is already open to that Neighbor. This suggests that as far as NUD is concerned, the Transient Neighbor is indistinguishable from an On-LL Neighbor.

3.4. Duplicate Address Detection.

Duplicate Address Detection is only required within the link-local scope, which in this case is the LL-local scope. Transient Neighbors are outside the scope of the LL. No particular interaction is required between the mechanism for establishing shortcuts and the mechanism for detection of duplicate link local addresses.

4 Node Operation Concepts.

This section describes node operations for performing basic functions (such as sending and receiving data) on a Logical Link. The application of these basic functions to the operation of the various IPv6 protocols such as Neighbor Discovery is described in Appendix A.

The majority of this section applies only to NBMA networks when used to provide point to point and point to multipoint SVCs. Section 7 discusses the case where the NBMA network is being used to supply only point to point PVCs.

4.1.Connecting to a Logical Link.

Before a node can send or receive IPv6 datagrams its underlying IPv6/NBMA interface(s) must first join a Logical Link.

An IPv6/NBMA driver SHALL establish a pt-pt VC to the MARS associated with its Logical Link, and register as a Cluster Member [5]. The node’s IPv6/NBMA interface will then be a member of the LL, have a Cluster Member ID (CMI) assigned, and can begin supporting IPv6 and IPv6 ND operations.

If the node is a host or router starting up it SHALL issue a single group MARS_JOIN for the following groups:

- Its derived Solicited-node address(es) with link-local scope.
- The All-nodes address with link-local scope.
- Other configured multicast groups with at least link-local scope.

If the node is a router it SHALL additionally issue:

- A single group MARS_JOIN for the All-routers address with link-local scope.
- A block MARS_JOIN for the range(s) of IPv6 multicast addresses (with greater than link-local scope) for which promiscuous reception is required.

The encapsulation mechanism for, and key field values of, MARS control messages SHALL be defined in companion documents specific to particular NBMA network technologies.

4.2 Joining a Multicast Group.

This section describes the node’s behavior when it gets a JoinLocalGroup request from the IPv6 Layer. The details of how this behavior is achieved are going to be implementation specific.

If a JoinLocalGroup for a node-local address is received, the IPv6/NBMA driver SHALL return success indication to the caller and take no additional action. (Packets sent to node-local addresses never reach the IPv6/NBMA driver.)

If a JoinLocalGroup is received for an address with greater than node-local scope, the IPv6/NBMA driver SHALL send an appropriate single group MARS_JOIN request to register this address with the MARS.

4.3. Leaving a Multicast Group.

This section describes the node's behavior when it gets a LeaveLocalGroup request from the IPv6 Layer. The details of how this behavior is achieved are going to be implementation specific.

If a LeaveLocalGroup for a node-local address is received, the IPv6/NBMA driver SHALL return success indication to the caller and take no additional action. (Packets sent to node-local addresses never reach the IPv6/NBMA driver.)

If a LeaveLocalGroup is received for an address with greater than node-local scope, the IPv6/NBMA driver SHALL send an appropriate single group MARS_LEAVE request to deregister this address with the MARS.

4.4. Sending Data.

Separate processing and encapsulation rules apply for outbound unicast and multicast packets.

4.4.1. Sending Unicast Data.

The IP level 'next hop' for each outbound unicast IPv6 packet is used to identify a pt-pt VC on which to forward the packet.

For NBMA networks where LLC/SNAP encapsulation is typically used (e.g. ATM or SMDS), the IPv6 packet SHALL be encapsulated with the following LLC/SNAP header and sent over the VC.

```
[0xAA-AA-03][0x00-00-00][0x86-DD][IPv6 packet]
    (LLC)       (OUI)     (PID)
```

For NBMA networks that do not use LLC/SNAP encapsulation, an alternative rule SHALL be specified in the NBMA-specific companion document.

If no pt-pt VC exists for the next hop address for the packet, the node SHALL place a call to set up a VC to the next hop destination. Any time the IPv6/NBMA driver receives a unicast packet for transmission the IPv6 layer will already have determined the link-layer (NBMA) address of the next hop. Thus, the information needed to place the NBMA call to the next hop will be available.

The sending node SHOULD queue the packet that triggered the call request, and send it when the call is established.

If the call to the next hop destination node fails the sending node SHALL discard the packet that triggered the call setup. Persistent failure to create a VC to the next hop destination will be detected and handled at the IPv6 Network Layer through NUD.

At this time no rules are specified for mapping outbound packets to VCs using anything more than the packet's destination address.

4.4.2. Sending Multicast Data.

The IP level 'next hop' for each outbound multicast IPv6 packet is used to identify a pt-pt or pt-mpt VC on which to forward the packet.

For NBMA networks where LLC/SNAP encapsulation is typically used (e.g. ATM or SMDS), multicast packets SHALL be encapsulated in the following manner:

```
[0xAA-AA-03][0x00-00-5E][0x00-01][pkt$cmi][0x86DD][IPv6
packet]
    (LLC)       (OUI)     (PID)    (mars encaps)
```

The IPv6/NBMA driver's Cluster Member ID SHALL be copied into the 2 octet pkt$cmi field prior to transmission.

For NBMA networks that do not use LLC/SNAP encapsulation, an alternative rule SHALL be specified in the NBMA-specific companion document. Some mechanism for carrying the IPv6/NBMA driver's Cluster Member ID SHALL be provided.

If the packet's destination is one of the following multicast addresses, it SHALL be sent over the IPv6/NBMA driver's direct pt-pt VC to the MARS:

- A Solicited-node address with link-local scope.
- The All-nodes address with link-local scope.
- The All-routers address with link-local scope.
- A DHCP-v6 relay or server multicast address.

The MARS SHALL then redistribute the IPv6 packet as described in section 3.1.1. (If the VC to the MARS has been idle timed out for some reason, it MUST be re-established before forwarding the packet to the MARS.)

If packet's destination is any other address, then the usual MARS client mechanisms are used by the IPv6/NBMA driver to select and/or establish a pt-mpt VC on which the packet is to be sent.

At this time no rules are specified for mapping outbound packets to VCs using anything more than the packet's destination address.

4.5. Receiving Data.

Packets received using the encapsulation shown in section 4.4.1 SHALL be de-encapsulated and passed up to the IPv6 layer. The IPv6 layer then determines how the incoming packet is to be handled.

Packets received using the encapsulation specified in section 4.4.2 SHALL have their pkt$cmi field compared to the local IPv6/NBMA driver's own CMI. If the pkt$cmi in the header matches the local CMI the packet SHALL be silently dropped. Otherwise, the packet SHALL be de-encapsulated and passed to the IPv6 layer. The IPv6 layer then determines how the incoming packet is to be handled.

For NBMA networks that do not use LLC/SNAP encapsulation, alternative rules SHALL be specified in the NBMA-specific companion document.

The IPv6/NBMA driver SHALL NOT attempt to filter out multicast IPv6 packets arriving with encapsulation defined for unicast packets, nor attempt to filter out unicast IPv6 packets arriving with encapsulation defined for multicast packets.

4.6. VC Setup and release for unicast data.

Unicast VCs are maintained separately from multicast VCs. The setup and maintenance of multicast VCs are handled by the MARS client in each IPv6/NBMA driver [5]. Only the setup and maintenance of pt-pt VCs for unicast IPv6 traffic will be described here. Only best effort unicast VCs are considered. The creation of VCs for other classes of service is outside the scope of this document.

Before sending a packet to a new destination within the same LL a node will first perform a Neighbor Discovery on the intra-LL target. This is done to resolve the IPv6 destination address into a link-layer address which the sender can then use to send unicast packets.

Appendix A.1.1 contains non-normative, descriptive text covering the Neighbor Solicitation/Advertisement exchange and eventual establishment of a new SVC.

A Redirect message (either a redirect to a node on the same LL, or a shortcut redirect to a node outside the LL) results in the sending (redirected) node creating a new pt-pt VC to a new receiving node. the Redirect message SHALL contain the link layer (NBMA) address of the new receiving IPv6/NBMA interface. The redirected node does not concern itself where the new receiving node is located on the NBMA network. The redirected node will set up a pt-pt VC to the new node if one does not previously exist. The redirected node will then use the new VC to send data rather than whatever VC it had previously been using.

Redirects are unidirectional. Even after the source has reacted to a redirect, the destination will continue to send IPv6 packets back to the redirected node on the old path. This happens because the destination node has no way of determining the IPv6 address of the other end of a new VC in the absence of Neighbor Discovery. Thus, redirects will not result in both ends of a connection using the new VC. IPv6 redirects are not intended to provide symmetrical redirection. If the non-redirected node eventually receives a redirect it MAY discover the existing VC to the target node and use that rather than creating a new VC.

It is desirable that VCs are released when no longer needed.

An IPv6/NBMA driver SHALL release any VC that has been idle for 20 minutes.

This time limit MAY be reduced through configuration or as specified in companion documents for specific NBMA networks.

If a Neighbor or Destination cache entry is purged then any VCs associated with the purged entry SHOULD be released.

If the state of an entry in the Neighbor cache is set to STALE, then any VCs associated with the stale entry SHOULD be released.

4.7 NBMA SVC Signaling Support and MTU issues.

Mechanisms for signaling the establishment and teardown of pt-pt and pt-mpt SVCs for different NBMA networks SHALL be specified in companion documents.

Since any given IPv6/NBMA driver will not know if the remote end of a VC is in the same LL, drivers SHALL implement NBMA-specific mechanisms to negotiate acceptable MTUs at the VC level. These mechanisms SHALL be specified in companion documents.

However, IPv6/NBMA drivers can assume that they will always be talking to another driver attached to the same type of NBMA network. (For example, an IPv6/NBMA driver does not need to consider the possibility of establishing a shortcut VC directly to an IPv6/FR driver.)

5. Interface Tokens, Link Layer Address Options, Link-Local Addresses

5.1 Interface Tokens

Each IPv6 interface must have an interface token from which to form IPv6 autoconfigured addresses. This interface token must be unique within a Logical Link to prevent the creation of duplicate addresses when stateless address configuration is used.

In cases where two nodes on the same LL produce the same interface token then one interface MUST choose another host-token. All implementations MUST support manual configuration of interface tokens to allow operators to manually change a interface token on a per-LL basis. Operators may choose to manually set interface tokens for reasons other than eliminating duplicate addresses.

All interface tokens MUST be 64 bits in length and formatted as described in the following sections. The hosts tokens will be based on the format of an EUI-64 identifier [10]. Refer to [19 - Appendix A] for a description of creating IPv6 EUI-64 based interface identifiers.

5.1.1 Single Logical Links on a Single NBMA Interface

Physical NBMA interfaces will generally have some local identifier that may be used to generate a unique IPv6/NBMA interface token. The exact mechanism for generating interface tokens SHALL be specified in companion documents specific to each NBMA network.

5.1.2 Multiple Logical Links on a Single NBMA Interface

Physical NBMA interfaces MAY be used to provide multiple logical NBMA interfaces. Since each logical NBMA interface MAY support an independent IPv6 interface, two separate scenarios are possible:

- A single host with separate IPv6/NBMA interfaces onto a number of independent Logical Links.

- A set of 2 or more 'virtual hosts' (vhosts) sharing a common NBMA driver. Each vhost is free to establish IPv6/NBMA interfaces associated with different or common LLs. However, vhosts are bound by the same requirement as normal hosts - no two interfaces to the same LL can share the same interface token.

In the first scenario, since each IPv6/NBMA interface is associated with a different LL, each interface's external identity can be differentiated by the LL's routing prefix. Thus, the host can re-use a single unique interface token across all its IPv6/NBMA interfaces. (Internally the host will tag received packets in some locally specific manner to identify what IPv6/NBMA interface they arrived on. However, this is an issue generic to IPv6, and does not required clarification in this document.)

The second scenario is more complex, but likely to be rarer.

When supporting multiple logical NBMA interfaces over a single physical NBMA interface, independent and unique identifiers SHALL be generated for each virtual NBMA interface to enable the construction of unique IPv6/NBMA interface tokens. The exact mechanism for generating interface tokens SHALL be specified in companion documents specific to each NBMA network.

5.2 Link Layer Address Options

Neighbor Discovery defines two option fields for carrying link-layer specific source and target addresses.

Between IPv6/NBMA interfaces, the format for these two options is adapted from the MARS [5] and NHRP [8] specs. It SHALL be:

```
[Type][Length][NTL][STL][..NBMA Number..][..NBMA
Subaddress..]
|   Fixed    ||            Link layer address
|
```

[Type] is a one octet field.

1 for Source link-layer address.
2 for Target link-layer address.

[Length] is a one octet field.

The total length of the option in multiples of 8 octets. Zeroed bytes are added to the end of the option to ensure its length is a multiple of 8 octets.

[NTL] is a one octet 'Number Type & Length' field.

[STL] is a one octet 'SubAddress Type & Length' field.

[NBMA Number] is a variable length field. It is always present. This contains the primary NBMA address.

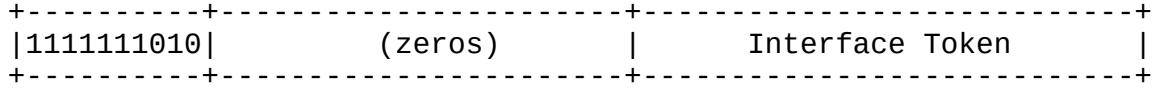
[NBMA Subaddress] is a variable length field. It may or may not be present. This contains any NBMA subaddress that may be required.

If the [NBMA Subaddress] is not present, the option ends after the [NBMA Number] (and any additional padding for 8 byte alignment).

The contents and interpretation of the [NTL], [STL], [NBMA Number], and [NBMA Subaddress] fields are specific to each NBMA network, and SHALL be specified in companion documents.

5.3 Link-Local Addresses

The IPv6 link-local address is formed by appending the interface token, as defined above, to the prefix FE80::/64.



6. Conclusion and Open Issues

This document describes a general architecture for IPv6 over NBMA networks. It forms the basis for subsidiary companion documents that provide details for various specific NBMA technologies (such as ATM or Frame Relay). The IPv6 over NBMA architecture allows conventional host-side operation of the IPv6 Neighbor Discovery protocol, while also supporting the establishment of 'shortcut' NBMA forwarding paths (when dynamically signaled NBMA links are available).

The IPv6 "Link" is generalized to "Logical Link" in an analagous manner to the IPv4 "Logical IP Subnet". The MARS protocol is augmented and used to provide relatively efficient intra Logical Link multicasting of IPv6 packets, and distribution of Discovery messages. Shortcut NBMA level paths are supported either through router based flow detection, or host originated explicit requests. Neighbor Discovery is used without modification for all intra-LL control (including the initiation of NBMA shortcut discovery). Router to router NHRP is used to obtain the IPv6/NBMA address mappings for shortcut targets outside a source's Logical Link.

7. Security Considerations

This architecture introduces no new protocols, but depends on existing protocols (NHRP, IPv6, ND, MARS) and is therefore subject to all the security threats inherent in these protocols. This architecture should not be used in a domain where any of the base protocols are considered unacceptably insecure. However, this protocol itself does not introduce additional security threats.

While this proposal does not introduce any new security mechanisms all current IPv6 security mechanisms will work without modification for NBMA. This includes both authentication and encryption for both Neighbor Discovery protocols as well as the exchange of IPv6 data packets. The MARS protocol is modified in a manner that does not affect or augment the security offered by RFC 2022.

Acknowledgments

Eric Nordmark confirmed the usefulness of ND Redirect messages in private email during the March 1996 IETF. The discussions with various ION WG members during the June and December 1996 IETF helped solidify the architecture described here. Grenville Armitage's original work on IPv6/NBMA occurred while employed at Bellcore. Elements of section 5 were borrowed from Matt Crawford's memo on IPv6 over Ethernet.

Authors' Addresses

Grenville Armitage
Bell Laboratories, Lucent Technologies
101 Crawfords Corner Road
Holmdel, NJ 07733
USA

EMail: gja@lucent.com

Peter Schulter
Bright Tiger Technologies
125 Nagog Park
Acton, MA 01720

EMail: paschulter@acm.org

Markus Jork
European Applied Research Center
Digital Equipment GmbH
CEC Karlsruhe
Vincenz-Priessnitz-Str. 1
D-76131 Karlsruhe
Germany

EMail: jork@kar.dec.com

Geraldine Harter
Digital UNIX Networking
Compaq Computer Corporation
110 Spit Brook Road
Nashua, NH 03062

EMail: harter@zk3.dec.com

References

[1] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[2] ATM Forum, "ATM User Network Interface (UNI) Specification Version 3.1", ISBN 0-13-393828-X, Prentice Hall, Englewood Cliffs, NJ, June 1995.

[3] Crawford, M., "A Method for the Transmission of IPv6 Packets over Ethernet Networks", RFC 1972, August 1996.

[4] Heinanen, J., "Multiprotocol Encapsulation over ATM Adaptation Layer 5", RFC 1483, July 1993.

[5] Armitage, G., "Support for Multicast over UNI 3.1 based ATM Networks", RFC 2022, November 1996.

[6] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 2373, July 1998.

[7] Narten, T., Nordmark, E. and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC 2461, December 1998.

[8] Luciani, J., Katz, D., Piscitello, D. Cole B and N. Doraswamy, "NBMA Next Hop Resolution Protocol (NHRP)", RFC 2332, April 1998.

[9] Thomson, S. and T. Narten, "IPv6 Stateless Address Autoconfiguration", RFC 2462, December 1998.

[10] "64-Bit Global Identifier Format Tutorial", http://standards.ieee.org/db/oui/tutorials/EUI64.html.

[11] Katsube, Y., Nagami, K. and H. Esaki, "Toshiba's Router Architecture Extensions for ATM : Overview", RFC 2098, February 1997.

[12] P. Newman, T. Lyon, G. Minshall, "Flow Labeled IP: ATM under IP", Proceedings of INFOCOM'96, San Francisco, March 1996, pp.1251-1260

[13] Piscitello, D. and J. Lawrence, "The Transmission of IP Datagrams over the SMDS Service", RFC 1209, March 1991.

[14] Plummer, D., "An Ethernet Address Resolution Protocol - or - Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware", STD 37, RFC 826, November 1982.

[15] McCann, J., Deering, S. and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

[16] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[17] Armitage, G., Schulter, P. and M. Jork, "IPv6 over ATM Networks", RFC 2492, January 1999.

[18] C. Perkins, J. Bound, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", Work in Progress.

[19] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 2373, July 1998.

Appendix A. IPv6 Protocol Operation Description

The IPv6 over NBMA model described in this document maintains the complete semantics of the IPv6 protocols. No changes need to be made to the IPv6 Network Layer. Since the concept of the security association is not being changed for NBMA, this framework maintains complete IPv6 security semantics and features. This allows IPv6 nodes to choose their responses to solicitations based on security information as is done with other datalinks, thereby maintaining the semantics of Neighbor Discovery since it is always the solicited node that chooses what (and even if) to reply to the solicitation. Thus, NBMA will be transparent to the network layer except in cases where extra services (such as QoS VCs) are offered.

The remainder of this Appendix describes how the core IPv6 protocols will operate within the model described here.

A.1 Neighbor Discovery Operations

Before performing any sort of Neighbor discover operation, each node must first join the all-node multicast group, and it’s solicited node multicast address (the use of this address in relation to DAD is described in A.1.4). The IPv6 network layer will join these multicast groups as described in 4.2.

A.1.1 Performing Address Resolution

An IPv6 host performs address resolution by sending a Neighbor Solicitation to the solicited-node multicast address of the target host, as described in [7]. The Neighbor Solicitation message will contain a Source Link-Layer Address Option set to the soliciting node’s NBMA address on the LL.

When the local node’s IPv6/NBMA driver is passed the Neighbor Solicitation message from the IPv6 network layer, it follows the steps described in section 4.4.2 Sending Multicast Data.

One or more nodes will receive the Neighbor Solicitation message. The nodes will process the data as described in section 4.5 and pass the de-encapsulated packets to the IPv6 network layer.

If the receiving node is the target of the Neighbor Solicitation it will update its Neighbor cache with the soliciting node’s NBMA address, contained in the Neighbor Solicitation message’s Source Link-Layer Address Option as described in [7].

The solicited IPv6 host will respond to the Neighbor Solicitation with a Neighbor Advertisement message sent to the IPv6 unicast address of the soliciting node. The Neighbor Advertisement message will contain a Target Link-Layer Address Option set to the solicited node's NBMA address on the LL.

The solicited node's IPv6/NBMA driver will be passed the Neighbor Advertisement and the soliciting node's link-layer address from the IPv6 network layer. It will then follow the steps described in section section 4.4.1 to send the NA message to the soliciting node. This will create a pt-pt VC between the solicited node and soliciting node if one did not already exist.

The soliciting node will then receive the Neighbor Advertisement message over the new PtP VC, de-encapsulate the message, and pass it to the IPv6 Network layer for processing as described in section 4.5. The soliciting node will then make the appropriate entries in it's Neighbor cache, including caching the NBMA link-layer address of the solicited node as described in [7].

At this point each system has a complete Neighbor cache entry for the other system. They can exchange data over the pt-pt VC newly created by the solicited node when it returned the Neighbor Advertisement, or create a new VC.

An IPv6 host can also send an Unsolicited Neighbor Advertisemnent to the all-nodes multicast address. When the local node IPv6/NBMA driver is passed the Neighbor Advertisement from the IPv6 network layer, it follows the steps described in section 4.4.2 to send the NA message to the all-nodes multicast address. Each node will process the incoming packet as described in section 4.5 and then pass the packet to the IPv6 network layer where it will be processed as described in [7].

A.1.2 Performing Router Discovery

Router Discovery is described in [7]. To support Router Discovery an IPv6 router will join the IPv6 all-routers multicast group address. When the IPv6/NBMA driver gets the JoinLocalGroup request from the IPv6 Network Layer, it follows the process described in section 4.2.

IPv6 routers periodically send unsolicited Router Advertisements announcing their availability on the LL. When an IPv6 router sends an unsolicited Router Advertisement, it sends a data packet addressed to the IPv6 all-nodes multicast address. When the local node IPv6/NBMA driver gets the Router Advertisement message from the IPv6 network layer, it transmits the message by following steps described in section 4.4.2. The MARS will transmit the packet on the LL's

ClusterControlVC, which sends the packets to all nodes on the LL. Each node on the LL will then process the incoming packet as described in section 4.5 and pass the received packet to the IPv6 Network layer for processing as appropriate.

To perform Router Discovery, an IPv6 host sends a Router Solicitation message to the all-routers multicast address. When the local node IPv6/NBMA driver gets the request from the IPv6 Network Layer to send the packet, it follows the steps described in section 4.4.2. The RS message will be sent to either those nodes which have joined the all-routers multicast group or to all nodes. The nodes which receive the RA message will process the message as described in section 4.5 and pass the RA message up to the IPv6 layer for processing. Only those nodes which are routers will process the message and respond to it.

An IPv6 router responds to a Router Solicitation by sending a Router Advertisement addressed to the IPv6 all-nodes multicast address if the source address of the Router Solicitation was the unspecified address. If the source address in the Router Solicitation is not the unspecified address, the the router will unicast the Router Advertisement to the soliciting node. If the router sends the Router Advertisement to the all-nodes multicast address then it follows the steps described above for unsolicited Router Advertisements.

If the Router Advertisement is to be unicast to the soliciting node, the IPv6 network layer will give the node’s IPv6/NBMA driver the Router Advertisement and link-layer address of the soliciting node (obtained through Address Resolution if necessary) which will send the packet according to the steps described in section 4.4.1 This will result in a new pt-pt VC being created between the router and the soliciting node if one did not already exist.

The soliciting node will receive and process the Router Advertisement as described in section 4.5 and will pass the RA message to the IPv6 network layer. The IPv6 network layer may, depending on the state of the Neighbor cache entry, update the Neighbor cache with the router’s NBMA address, contained in the Router Advertisement message’s Source Link-Layer Address Option.

If a pt-pt VC is set up during Router Discovery, subsequent IPv6 best effort unicast data between the soliciting node and the router will be transmitted over the new PtP VC.

A.1.3 Performing Neighbor Unreachability Detection (NUD)

Neighbor Unreachability Detection (NUD) is the process by which an IPv6 host determines that a neighbor is no longer reachable, as described in [7]. Each Neighbor cache entry contains information used by the NUD algorithm to detect reachability failures. Confirmation of a neighbor's reachability comes either from upper-layer protocol indications that data recently sent to the neighbor was received, or from the receipt of a Neighbor Advertisement message in response to a Neighbor Solicitation probe.

Connectivity failures at the node's IPv6/NBMA driver, such as released VCs (see section 4.6) and the inability to create a VC to a neighbor (see section 4.4.1), are detected and handled at the IPv6 network layer, through Neighbor Unreachability Detection. The node's IPv6/NBMA driver does not attempt to detect or recover from these conditions.

A persistent failure to create a VC from the IPv6 host to one of its IPv6 neighbors will be detected and handled through NUD. On each attempt to send data from the IPv6 host to its neighbor, the node's IPv6/NBMA driver will attempt to set up a VC to the neighbor, and failing to do so, will drop the packet. IPv6 reachability confirmation timers will eventually expire, and the neighbor's Neighbor cache entry will enter the PROBE state. The PROBE state will cause the IPv6 host to unicast Neighbor Solicitations to the neighbor, which will be dropped by the local node's IPv6/NBMA driver after again failing to setup the VC. The IPv6 host will therefore never receive the solicited Neighbor Advertisements needed for reachability confirmation, causing the neighbor's entry to be deleted from the Neighbor cache. The next time the IPv6 host tries to send data to that neighbor, address resolution will be performed. Depending on the reason for the previous failure, connectivity to the neighbor could be re-established (for example, if the previous VC setup failure was caused by an obsolete link-layer address in the Neighbor cache).

In the event that a VC from an IPv6 neighbor is released, the next time a packet is sent from the IPv6 host to the neighbor, the node's IPv6/NBMA driver will recognize that it no longer has a VC to that neighbor and attempt to setup a new VC to the neighbor. If, on the first and on subsequent transmissions, the node is unable to create a VC to the neighbor, NUD will detect and handle the failure as described earlier (handling the persistent failure to create a VC from the IPv6 host to one of its IPv6 neighbors). Depending on the reason for the previous failure, connectivity to the neighbor may or may not be re-established.

A.1.4 Performing Duplicate Address Detection (DAD)

An IPv6 host performs Duplicate Address Detection (DAD) to determine that the address it wishes to use on the LL (i.e. a tentative address) is not already in use, as described in [9] and [7]. Duplicate Address Detection is performed on all addresses the host wishes to use, regardless of the configuration mechanism used to obtain the address.

Prior to performing Duplicate Address Detection, a host will join the all-nodes multicast address and the solicited-node multicast address corresponding to the host's tentative address (see 4.2. Joining a Multicast Group). The IPv6 host initiates Duplicate Address Detection by sending a Neighbor Solicitation to solicited-node multicast address corresponding to the host's tentative address, with the tentative address as the target. When the local node's IPv6/NBMA driver gets the Neighbor Solicitation message from the IPv6 network layer, it follows the steps outlined in section 4.4.2. The NS message will be sent to those nodes which joined the target solicited-node multicast group or to all nodes. The DAD NS message will be received by one or more nodes on the LL and processed by each as described in section 4.5. Note that the MARS client of the sending node will filter out the message so that the sending node's IPv6 network layer will not see the message. The IPv6 network layer of any node which is not a member of the target solicited-node multicast group will discard the Neighbor Solicitation message.

If no other hosts have joined the solicited-node multicast address corresponding to the tentative address, then the host will not receive a Neighbor Advertisement containing its tentative address as the target. The host will perform the retransmission logic described in [9], terminate Duplicate Address Detection, and assign the tentative address to the NBMA interface.

Otherwise, other hosts on the LL that have joined the solicited-node multicast address corresponding to the tentative address will process the Neighbor Solicitation. The processing will depend on whether or not receiving IPv6 host considers the target address to be tentative.

If the receiving IPv6 host's address is not tentative, the host will respond with a Neighbor Advertisement containing the target address. Because the source of the Neighbor Solicitation is the unspecified address, the host sends the Neighbor Advertisement to the all-nodes multicast address following the steps outlined in section 4.4.2. The DAD NA message will be received and processed by the MARS clients on all nodes in the LL as described in section 4.5. Note that the sending node will filter the incoming message since the CMI in the message header will match that of the receiving node. All other

nodes will de-encapsulate the message and pass it to the IPv6 network layer. The host performing DAD will detect that its tentative address is the target of the Neighbor Advertisement, and determine that the tentative address is not unique and cannot be assigned to its NBMA interface.

If the receiving IPv6 host's address is tentative, then both hosts are performing DAD using the same tentative address. The receiving host will determine that the tentative address is not unique and cannot be assigned to its NBMA interface.

A.1.5 Processing Redirects

An IPv6 router uses a Redirect Message to inform an IPv6 host of a better first-hop for reaching a particular destination, as described in [7]. This can be used to direct hosts to a better first hop router, another host on the same LL, or to a transient neighbor on another LL. The IPv6 router will unicast the Redirect to the IPv6 source address that triggered the Redirect. The router's IPv6/NBMA driver will transmit the Redirect message using the procedure described in section 4.4.1. This will create a VC between the router and the redirected host if one did not previously exist.

The IPv6/NBMA driver of the IPv6 host that triggered the Redirect will receive the encapsulated Redirect over one of it's pt-pt VCs. It will the de-encapsulate the packet, and pass the Redirect message to the IPv6 Network Layer, as described section 4.5.

Subsequent data sent from the IPv6 host to the destination will be sent to the next-hop address specified in the Redirect Message. For NBMA networks, the Redirect Message should contain the link-layer address option as described in [7] and section 5.2, thus the redirected node will not have to perform a Neighbor Solicitation to learn the link-layer address of the node to which it has been redirected. Thus, the redirect can be to any node on the NBMA network, regardless of the LL membership of the new target node. This allows NBMA hosts to be redirected off their LL to achieve shortcut by using standard IPv6 protocols.

Once redirected, the IPv6 network layer will give the node's IPv6/NBMA driver the IPv6 packet and the link-layer address of the next-hop node when it sends data to the redirected destination. The node's IPv6/NBMA driver will determine if a VC to the next-hop destination exists. If a pt-pt VC does not exist, then the IPv6/NBMA driver will queue the data packet and initiate a setup of a VC to the destination. When the VC is created, or if one already exists, then the node will encapsulate the outgoing data packet and send it on the VC.

Note that Redirects are unidirectional. The redirected host will create a VC to the next-hop destination as specified in the Redirect message, but the next-hop will not be redirected to the source host. Because no Neighbor Discovery takes place, the next-hop destination has no way of determining the identity of the caller when it receives the new VC. Also, since ND does not take place on redirects, the next-hop receives no event that would cause it to update it's neighbor or destination caches. However, it will continue to transmit data back to the redirected host on the former path to the redirected host. The next-hop node should be able to use the new VC from the redirected destination if it too receives a redirect redirecting it to the redirected node. This behavior is consistent with [7].

A.2 Address Configuration

IPv6 addresses are auto-configured using the stateless or stateful address auto-configuration mechanisms, as described in [9] and [18]. The IPv6 auto-configuration process involves creating and verifying the uniqueness of a link-local address on an LL, determining whether to use stateless and/or stateful configurationmechanisms to obtain addresses, and determining if other (non- address) information is to be autoconfigured. IPv6 addresses can also be manually configured, if for example, auto-configuration fails because the autoconfigured link-local address is not unique. An LL administrator specifies the type of autoconfiguration to use; the hosts on an LL receive this autoconfiguration information through Router Advertisement messages.

The following sections describe how stateless, stateful and manual address configuration will work in an IPv6/NBMA environment.

A.2.1 Stateless Address Configuration

IPv6 stateless address configuration is the process by which an IPv6 host autoconfigures its interfaces, as described in [IPV6-ADDRCONF].

When an IPv6 host first starts up, it generates a link-local address for the interface attached to the Logical Link. It then verifies the uniqueness of the link-local address using Duplicate Address Detection (DAD). If the IPv6 host detects that the link-local address is not unique, the autoconfiguration process terminates. The IPv6 host must then be manually configured.

After the IPv6 host determines that the link-local address is unique and has assigned it to the interface on the Logical Link, the IPv6 host will perform Router Discovery to obtain auto-configuration information. The IPv6 host will send out a Router Solicitation and will receive a Router Advertisement, or it will wait for an

unsolicited Router Advertisement. The IPv6 host will process the M and O bits of the Router Advertisement, as described in [9] and as a result may invoke stateful address auto- configuration.

If there are no routers on the Logical Link, the IPv6 host will be able to communicate with other IPv6 hosts on the Logical Link using link-local addresses. The IPv6 host will obtain a neighbor's link-layer address using Address Resolution. The IPv6 host will also attempt to invoke stateful auto-configuration, unless it has been explicitly configured not to do so.

A.2.2 Stateful Address Configuration (DHCP)

IPv6 hosts use the Dynamic Host Configuration Protocol (DHCPv6) to perform stateful address auto-configuration, as described in [18].

A DHCPv6 server or relay agent is present on a Logical Link that has been configured with manual or stateful auto-configuration. The DHCPv6 server or relay agent will join the IPv6 DHCPv6 Server/Relay-Agent multicast group on the Logical Link. When the node's IPv6/NBMA driver gets the JoinLocalGroup request from the IPv6 network layer, it follows the process described in section 4.2.

An IPv6 host will invoke stateful auto-configuration if M and O bits of Router Advertisements indicate it should do so, and may invoke stateful auto-configuration if it detects that no routers are present on the Logical Link. An IPv6 host that is obtaining configuration information through the stateful mechanism will hereafter be referred to as a DHCPv6 client.

A DHCPv6 client will send a DHCPv6 Solicit message to the DHCPv6 Server/Relay-Agent multicast address to locate a DHCPv6 Agent. When the soliciting node's IPv6/NBMA driver gets the request from the IPv6 Network Layer to send the packet, it follows the steps described in section 4.4.2. This will result in one or more nodes on the LL receiving the message. Each node that receives the solicitation packet will process it as described in section section 4.5. Only the IPv6 network layer of the DHCPv6 server/relay-agent will accept the packet and process it.

A DHCPv6 Server or Relay Agent on the Logical Link will unicast a DHCPv6 Advertisement to the DHCPv6 client. The IPv6 network layer will give the node's IPv6/NBMA driver the packet and link-layer address of the DHCPv6 client (obtained through Neighbor Discovery if necessary). The node IPv6/NBMA driver will then transmit the packet as described in section 4.4.1. This will result in a new pt-pt VC being created between the server and the client if one did not previously exist.

The DHCP client's IPv6/NBMA driver will receive the encapsulated packet from the DHCP Server or Relay Agent, as described in section 4.5. The node will de-encapsulate the multicast packet and then pass it up to the IPv6 Network Layer for processing. The IPv6 network layer will deliver the DHCPv6 Advertise message to the DHCPv6 client.

Other DHCPv6 messages (Request, Reply, Release and Reconfigure) are unicast between the DHCPv6 client and the DHCPv6 Server. Depending on the reachability of the DHCPv6 client's address, messages exchanged between a DHCPv6 client and a DHCPv6 Server on another LL are sent either via a router or DHCPv6 Relay-Agent. Prior to sending the DHCPv6 message, the IPv6 network layer will perform Neighbor Discovery (if necessary) to obtain the link-layer address corresponding to the packet's next-hop. A pt-pt VC will be set up between the sender and the next hop, and the encapsulated packet transmitted over it, as described in 4.4. Sending Data.

A.2.3 Manual Address Configuration

An IPv6 host will be manually configured if it discovers through DAD that its link-local address is not unique. Once the IPv6 host is configured with a unique interface token, the auto-configuration mechanisms can then be invoked.

A.3 Internet Group Management Protocol (IGMP)

IPv6 multicast routers will use the IGMPv6 protocol to periodically determine group memberships of local hosts. In the framework described here, the IGMPv6 protocols can be used without any special modifications for NBMA. While these protocols might not be the most efficient in this environment, they will still work as described below. However, IPv6 multicast routers connected to an NBMA LL could optionally optimize the IGMP functions by sending MARS_GROUPLIST_REQUEST messages to the MARS serving the LL and determining group memberships by the MARS_GROUPLIST_REPLY messages. Querying the MARS for multicast group membership is an optional enchancement and is not required for routers to determine IPv6 multicast group membership on a LL.

There are three ICMPv6 message types that carry multicast group membership information: the Group Membership Query, Group Membership Report and Group Membership Reduction messages. IGMPv6 will continue to work unmodified over the IPv6/NBMA architecture described in this document.

An IPv6 multicast router receives all IPv6 multicast packets on the LL by joining all multicast groups in promiscuous mode [5]. The MARS server will then cause the multicast router to be added to all

existing and future multicast VCs. The IPv6 multicast router will thereafter be the recipient of all IPv6 multicast packets sent within the Logical Link.

An IPv6 multicast router discovers which multicast groups have members in the Logical Link by periodically sending Group Membership Query messages to the IPv6 all-nodes multicast address. When the local node's IPv6/NBMA driver gets the request from the IPv6 network layer to send the Group Membership Query packet, it follows the steps described in 4.4.2. The node determines that the destination address of the packet is the all-nodes multicast address and passes the packet to the node's MARS client where the packet is encapsulated and directly transmitted to the MARS. The MARS then relays the packet to all nodes in the LL. Each node's IPv6/NBMA drivers will receive the packet, de-encapsulate it, and passed it up to the IPv6 Network layer. If the originating node receives the encapsulated packet, the packet will be filtered out by the MARS client since the Cluster Member ID of the receiving node will match the CMI in the packet's MARS encapsulation header.

IPv6 hosts in the Logical Link will respond to a Group Membership Query with a Group Membership Report for each IPv6 multicast group joined by the host. IPv6 hosts can also transmit a Group Membership Report when the host joins a new IPv6 multicast group. The Group Membership Report is sent to the multicast group whose address is being reported. When the local node IPv6/NBMA driver gets the request from the IPv6 network layer to send the packet, it follows the steps described in 4.4.2. The node determines that the packet is being sent to a multicast address so forwards it to the node's MARS client for sending on the appropriate VC.

The Group Membership Report packets will arrive at every node which is a member of the group being reported through one of the VC attached to each node's MARS client. The MARS client will de-encapsulate the incoming packet and the packet will be passed to the IPv6 network layer for processing. The MARS client of the sending node will filter out the packet when it receives it.

An IPv6 host sends a Group Membership Reduction message when the host leaves an IPv6 multicast group. The Group Membership Reduction is sent to the multicast group the IPv6 host is leaving. The transmission and receipt of Group Membership Reduction messages are handled in the same manner as Group Membership Reports.

Appendix B. Alternative models of MARS support for Intra-LL ND

B.1 Simplistic approach - Use MARS 'as is'.

The IPv6/NBMA driver utilizes the standard MARS protocol to establish a VC forwarding path out of the interface on which it can transmit all multicast IPv6 packets, including ICMPv6 packets. The IPv6 packets are then transmitted, and received by the intended destination set, using separate pt-mpt VCs per destination group.

In this approach all the protocol elements in [5] are used 'as is'. However, SVC resource consumption must be taken into consideration. Unfortunately, ND assumes that link level multicast resources are best conserved by generating a sparsely distributed set of Solicited Node multicast addresses (to which discovery queries are initially sent). The original goal was to minimize the number of innocent nodes that simultaneously received discovery messages really intended for someone else.

However, in connection oriented NBMA environments it becomes equally (or more) important to minimize the number of independent VCs that a given NBMA interface is required to originate or terminate. If we treat the MARS service as a 'black box' the sparse Solicited Node address space can lead to a large number of short-use, but longer lived, pt-mpt VCs (generated whenever the node is transmitting Neighbor Solicitations). Even more annoying, these VCs are only useful for additional packets being sent to their associated Solicited Node multicast address. A new pt-pt VC is required to actually carry the unicast IPv6 traffic that prompted the Neighbor Solicitation.

The axis of inefficiency brought about by the sparse Solicited Nodes address space is orthogonal to the VC mesh vs Multicast Server tradeoff. Typically a multicast server aggregates traffic flow to a common multicast group onto a single VC. To reduce the VC consumption for ND, we need to aggregate across the Solicited Node address space - performing aggregation on the basis of a packet's function rather than its explicit IPv6 destination. The trade-off here is that the aggregation removes the original value of scattering nodes sparsely across the Solicited Nodes space. This is a price of the mismatch between ND and connection oriented networks.

B.2 MARS as a Link (Multicast) Server.

One possible aggregation mechanism is for every node's IPv6/NBMA driver to trap multicast ICMPv6 packets carrying multicast ND or RD messages, and logically remap their destinations to the All Nodes

group (link local scope). By ensuring that the All Nodes group is supported by an MCS, the resultant VC load within the LL will be significantly reduced.

A further optimization is for every node's IPv6/NBMA driver to trap multicast ICMPv6 packets carrying multicast ND or RD messages, and send them to the MARS itself for retransmission on ClusterControlVC (involving a trivial extension to the MARS itself.) This approach recognizes that in any LL where IPv6 multicasting is supported:

- Nodes already have a pt-pt VC to their MARS.

- The MARS has a pt-mpt VC (ClusterControlVC) out to all Cluster members (LL members registered for multicast support).

Because the VCs between a MARS and its MARS clients carry LLC/SNAP encapsulated packets, ICMP packets can be multiplexed along with normal MARS control messages. In essence the MARS behaves as a multicast server for non-MARS packets that it receives from around the LL.

As there is no requirement that a MARS client accepts only MARS control messages on ClusterControlVC, ICMP packets received in this fashion may be passed to every node's IP layer without further comment. Within the IP layer, filtering will occur based on the packet's actual destination IP address, and only the targeted node will end up responding.

Regrettably this approach does result in the entire Cluster's membership having to receive a variety of ICMPv6 messages that they will always throw away.

Appendix C. Flow detection

The relationship between IPv6 packet flows, Quality of Service guarantees, and optimal use of underlying IP and NBMA network resources are still subjects of ongoing research in the IETF (specifically the ISSLL, RSVP, IPNG, and ION working groups). This document currently only describes the use of flow detection as a means to optimize the use of NBMA network resources through the establishment of inter-LL shortcuts.

C.1. The use of non-zero FlowID to suppress flow detection

For the purposes of this IPv6/NBMA architecture, a flow is:

> A related sequence of IPv6 packets that the first hop router is allowed to perform flow-detection on for the purposes of triggering shortcut discovery.

How these packets are considered to be related to each other (e.g. through common header fields such as IPv6 destination addresses) is a local configuration issue.

The flow-detection rule specifies that only packets with a zero FlowID can be considered as flows for which shortcut discovery may be triggered. The rationale behind this decision is:

> NBMA shortcuts are for the benefit of ’the network’ optimizing its forwarding of IPv6 packets in the absence of any other guidance from the host.
>
> It is desirable for an IPv6/NBMA host to have some mechanism for overriding attempts by ’the network’ to optimize its internal forwarding path.
>
> A zero FlowID has IPv6 semantics of "the source allows the network to utilize its own discretion in providing best-effort forwarding service for packets with zero FlowID"
>
> The IPv6 semantics of zero FlowID are consistent with the flow-detection rule in this document of "if the FlowID is zero, we are free to optimize the forwarding path using shortcuts"
>
> A non-zero FlowID has IPv6 semantics of "the source has previously established some preferred, end to end hop by hop forwarding behaviour for packets with this FlowID"

The IPv6 semantics of non-zero FlowID are consistent with the flow-detection rule in this document of "if the FlowID is non-zero, do not attempt to impose a shortcut".

A non-zero FlowID might be assigned by the source host after negotiating a preferred forwarding mechanism with 'the network' (e.g. through dynamic means such as RSVP, or administrative means). Alternatively it can simply be assigned randomly by the source host, and the network will provide default best effort forwarding (an IPv6 router defaults to providing best-effort forwarding for packets whose FlowID/source-address pair is not recognized).

Thus, the modes of operation supported by this document becomes:

Zero FlowID
Best effort forwarding, with optional shortcut discovery triggered through flow-detection.

Non-zero FlowID
Best effort forwarding if the routers along the path have not been otherwise configured with alternative processing rules for this FlowID/source-address pair. Flow detection relating to shortcut discovery is suspended.

If the routers along the path have been configured with particular processing rules for this FlowID/source-address pair, the flow is handled according to those rules. Flow detection relating to shortcut discovery is suspended.

Mechanisms for establishing particular per-hop processing rules for packets with non-zero FlowID are neither constrained by, nor implied by, this document.

C.2. Future directions for Flow Detection

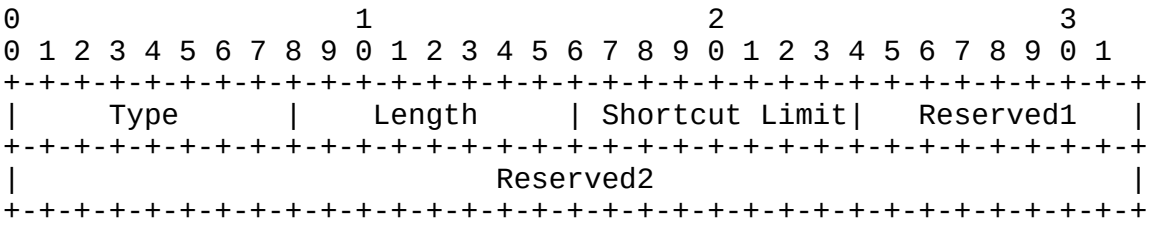
In the future, accurate mapping of IPv6 flows onto NBMA VCs may require more information to be exchanged during the Neighbor Discovery process than is currently available in Neighbor Discovery packets. In these cases, the IPv6 Neighbor Discover protocols can be extended to include new TLV options (see section 4.6 of RFC 1970 [7]). However, if new options are required, the specification of these options must be co-ordinated with the IPNG working group. Since RFC 1970 specifies that nodes must silently ignore options they do not understand, new options can be added at any time without breaking backward compatibility with existing implementations.

NHRP also provides mechanisms for adding optional TLVs to NHRP Requests and NHRP Replies. Future developments of this document's architecture will require consistent QoS extensions to both ND and NHRP in order to ensure they are semantically equivalent (syntactic differences are undesirable, but can be tolerated).

Support for QoS on IPv6 unicast flows will not require further extensions to the existing MARS protocol. However, future support for QoS on IPv6 multicast flows may require extensions. MARS control messages share the same TLV extension mechanism as NHRP, allowing QoS extensions to be developed as needed.

Appendix D. Shortcut Limit Option

For NS messages sent as a shortcut trigger, a new type of ND option is needed to pass on the information about the data flow hop limit from the host to the router. The use of this ND option is defined in section 3.2.2 of this specification. Its binary representation follows the rules of section 4.6 of RFC 1970:



Fields:

Type	6
Length	1
Shortcut Limit	8-bit unsigned integer. Hop limit for shortcut attempt.
Reserved1	This field is unused. It MUST be initialized to zero by the sender and MUST be ignored by the receiver.
Reserved2	This field is unused. It MUST be initialized to zero by the sender and MUST be ignored by the receiver.

Description

The shortcut limit option is used by a host in a Neighbor Solicitation message sent as a shortcut trigger to a default router. It restricts the router's shortcut query to targets reachable via the specified number of hops. The shortcut limit is given relative to the host requesting the shortcut. NS messages with shortcut limit values of 0 or 1 MUST be silently ignored.

Full Copyright Statement