Routing over Large Clouds Working Group
INTERNET-DRAFT
<draft-ietf-rolc-nhrp-04.txt>

Dave Katz
(cisco Systems)
David Piscitello
(Core Competence, Inc.)
May, 1995

NBMA Next Hop Resolution Protocol (NHRP)

Status of this Memo

Abstract

This document describes the NBMA Next Hop Resolution Protocol (NHRP). NHRP can be used by a source station (host or router) connected to a Non-Broadcast, Multi-Access (NBMA) subnetwork to determine the IP and NBMA subnetwork addresses of the "NBMA next hop" towards a destination station. If the destination is connected to the NBMA subnetwork, then the NBMA next hop is the destination station itself. Otherwise, the NBMA next hop is the egress router from the NBMA subnetwork that is "nearest" to the destination station. Although this document focuses on NHRP in the context of IP, the technique is applicable to other internetwork layer protocols (e.g., IPX, CLNP, Appletalk) as well.

This document is intended to be a functional superset of the NBMA Address Resolution Protocol (NARP) documented in [1].

Operation of NHRP as a means of establishing a transit path across an NBMA subnetwork between two routers will be addressed in a separate document.

1. Introduction

The NBMA Next Hop Resolution Protocol (NHRP) allows a source station (a host or router), wishing to communicate over a Non-Broadcast, Multi-Access (NBMA) subnetwork, to determine the IP and NBMA addresses of the "NBMA next hop" toward a destination station. A subnetwork can be non-broadcast either because it technically doesn't support broadcasting (e.g., an X.25 subnetwork) or because broadcasting is not feasible for one reason or another (e.g., an SMDS multicast group or an extended Ethernet would be too large). If the destination is connected to the NBMA subnetwork, then the NBMA next hop is the destination station itself. Otherwise, the NBMA next hop is the egress router from the NBMA subnetwork that is "nearest" to the destination station.

An NBMA subnetwork may, in general, consist of multiple logically independent IP subnets (LISs), defined in [3] and [4] as having the following properties:

1) All members of a LIS have the same IP network/subnet number and address mask.

2) All members within a LIS are directly connected to the same NBMA subnetwork.

3) All members outside of the LIS are accessed via a router.

IP routing described in [3] and [4] only resolves the next hop address if the destination station is a member of the same LIS as the source station; otherwise, the source station must forward packets to a router that is a member of multiple LIS's. In multi-LIS configurations, hop-by-hop IP routing may not be sufficient to resolve the "NBMA next hop" toward the destination station, and IP packets may traverse the NBMA subnetwork more than once.

NHRP describes a routing method that relaxes the forwarding restrictions of the LIS model. With NHRP, once the NBMA next hop has been resolved, the source may either start sending IP packets to the destination (in a connectionless NBMA subnetwork such as SMDS) or may first establish a connection to the destination with the desired bandwidth and QOS characteristics (in a connection-oriented NBMA subnetwork such as ATM).

NHRP in its most basic form provides a simple IP-to-NBMA-address binding service. This may be sufficient for hosts which are directly connected to an NBMA subnetwork, allowing for straightforward implementations in NBMA stations. NHRP also has the capability of determining the egress point from an NBMA subnetwork when the

destination is not directly connected to the NBMA subnetwork and the identity of the egress router is not learned by other methods (such as routing protocols). Optional extensions to NHRP provide additional robustness and diagnosability.

NHRP supports both a server-based style of deployment and a ubiquitous "fabric", consisting of NHRP-capable routers. The server-based approach requires a smaller number of machines (possibly one) to support NHRP, but requires significantly more manual configuration.

Address resolution techniques such as those described in [3] and [4] may be in use when NHRP is deployed. ARP servers and services over NBMA subnetworks may be required to support hosts that are not capable of dealing with any model for communication other than the LIS model, and deployed hosts may not implement NHRP but may continue to support ARP variants such as those described in [3] and [4]. NHRP is intended to reduce or eliminate the extra router hops required by the LIS model, and can be deployed in a non-interfering manner alongside existing ARP services.

The operation of NHRP to establish transit paths across NBMA subnetworks between two routers requires additional mechanisms to avoid stable routing loops, and will be described in a separate document.

2. Overview

2.1 Terminology

The term "network" is highly overloaded, and is especially confusing in the context of NHRP. We use the following terms:

Internetwork layer--the media-independent layer (IP in the case of TCP/IP networks).

Subnetwork layer--the media-dependent layer underlying the internetwork layer, including the NBMA technology (ATM, X.25, SMDS, etc.)

2.2 Protocol Overview

In this section, we briefly describe how a source S (which potentially can be either a router or a host) uses NHRP to determine the "NBMA next hop" to destination D.

For administrative and policy reasons, a physical NBMA subnetwork may be partitioned into several, disjoint "Logical NBMA subnetworks". A Logical NBMA subnetwork is defined as a collection of hosts and routers that share unfiltered subnetwork connectivity over an NBMA subnetwork. "Unfiltered subnetwork connectivity" refers to the absence of closed user groups, address screening or similar features that may be used to prevent direct communication between stations connected to the same NBMA subnetwork. (Hereafter, unless otherwise specified, we use the term "NBMA subnetwork" to mean *logical* NBMA subnetwork.)

Placed within the NBMA subnetwork are one or more entities that implement the NHRP protocol, otherwise known as "Next Hop Servers" (NHSs). Each NHS serves a set of destination hosts, which may or may not be directly connected to the NBMA subnetwork. NHSs cooperatively resolve the NBMA next hop within their logical NBMA subnetwork. In addition to NHRP, NHSs may participate in protocols used to disseminate routing information across (and beyond the boundaries of) the NBMA subnetwork, and may support "classical" ARP service as well.

An NHS maintains a "next-hop resolution" cache, which is a table of address mappings (IP-to-NBMA address). This table can be constructed from information gleaned from NHRP Register packets (see Section 5.4), extracted from NHRP requests or replies that traverse the NHS as they are forwarded, or through mechanisms outside the scope of this document (examples of such mechanisms include ARP [2, 3, 4] and pre-configured tables). Section 6.3 further describes cache management issues.

A host or router that is not an NHRP server must be configured with the identity of the NHS which serves it (see Configuration, Section 4).

[Note: for NBMA subnetworks that offer group or multicast addressing features, it may be desirable to configure stations with a group identity for NHSs, i.e., addressing information that would solicit a response from "all NHSs". The means whereby a group of NHSs divide responsibilities for next hop resolution are not described here.]

The protocol proceeds as follows. An event occurs triggering station S to want to resolve the NBMA address of a path to D. This is most likely to be when a data packet addressed to station D is to be emitted from station S (either because station S is a host, or station S is a transit router), but the address resolution could also be triggered by other means (a resource reservation request, for example). Station S first determines the next hop to station D through normal routing processes (for a host, the next hop may simply be the default router; for routers, this is the "next hop" to the

destination IP address). If the next hop is reachable through its NBMA interface, S constructs an NHRP request packet (see Section 5.2) containing station D's IP address as the (target) destination address, S's own IP address as the source address (NHRP request initiator), and station S's NBMA addressing information. Station S may also indicate that it prefers an authoritative reply (i.e., station S only wishes to receive a reply from the NHS-speaker that maintains the NBMA-to-IP address mapping for this destination). Station S encapsulates the NHRP request packet in an IP packet containing as its destination address the IP address of its configured NHS. This IP packet is emitted across the NBMA interface to the NBMA address of the NHS.

If the NHRP request is triggered by a data packet, station S may choose to dispose of the data packet while awaiting an NHRP reply in one of the following ways:

(a) Drop the packet
(b) Retain the packet until the reply arrives and a more optimal path is available
(c) Forward the packet along the routed path toward D

The choice of which of the above to perform is a local policy matter, though option (c) is the recommended default, since it may allow data to flow to the destination while the NBMA address is being resolved. Note that an NHRP request for a given destination MUST NOT be triggered on every packet, though periodically retrying a request is permitted.

When the NHS receives an NHRP request, a check is made to see if it "serves" station D, i.e., the NHS checks to see if there is a "next hop" entry for D in its next-hop resolution cache. If the NHS does not serve D, the NHS forwards the NHRP request to another NHS. (Mechanisms for determining how to forward the NHRP request are discussed in Section 3, Modes of Deployment.)

If this NHS serves D, the NHS resolves station D's NBMA address, and generates a positive NHRP reply on D's behalf. (NHRP replies in this scenario are always marked as "authoritative".) The NHRP reply packet contains the next hop IP and NBMA address for station D and is sent back to S. (Note that if station D is not on the NBMA subnetwork, the next hop IP address will be that of the egress router through which packets for station D are forwarded.)

An NHS receiving an NHRP reply may cache the NBMA next hop information contained therein. To a subsequent NHRP request, this NHS may respond with the cached, non-authoritative, NBMA next hop information or with cached negative information, or may not be

allowed to respond with the cached information (see section 6.3). Non-authoritative NHRP replies are distinguished from authoritative replies so that if a communication attempt based on non-authoritative information fails, a source station can choose to send an authoritative NHRP request. NHSs MUST NOT respond to authoritative NHRP requests with cached information.

[Note: An NHRP reply can be returned directly to the NHRP request initiator, i.e., without traversing the list of NHSs that forwarded the request, if all of the following criteria are satisfied:

(a) Direct communication is available via datagram transfer (e.g., SMDS) or the NHS has an existing virtual circuit connection to the NHRP request initiator or is permitted to open one.
(b) The NHRP request initiator has not included the NHRP Reverse NHS record Extension (see Section 5.7.5).
(c) The authentication policy in force permits direct communication between the NHS and the NHRP request initiator.

The purpose of allowing an NHS to reply directly is to reduce response time. A consequence of allowing a direct reply is that NHSs that would under normal circumstances be traversed by the reply would not cache next hop information contained therein.]

The process of forwarding the NHRP request is repeated until the request is satisfied, or an error occurs (e.g., no NHS in the NBMA subnetwork can resolve the request.) If the determination is made that station D's next hop cannot be resolved, a negative reply is returned. This occurs when (a) no next-hop resolution information is available for station D from any NHS, or (b) an NHS is unable to forward the NHRP request (e.g., connectivity is lost).

NHRP requests and replies MUST NOT cross the borders of a logical NBMA subnetwork (an explicit NBMA subnetwork identifier may be included as an extension in the NHRP request, see section 5.7.2). Thus, IP traffic out of and into a logical NBMA subnetwork always traverses an IP router at its border. Internetwork layer filtering can then be implemented at these border routers.

NHRP optionally provides a mechanism to reply with aggregated NBMA next hop information. Suppose that router X is the NBMA next hop from station S to station D. Suppose further that X is an egress router for all stations sharing an IP address prefix with station D. When an NHRP reply is generated in response to a request, the responder may augment the IP address of station D with a bit count defining this prefix (see Section 5.7.1). A subsequent (non-

authoritative) NHRP request for some destination that shares an IP address prefix with D may be satisfied with this cached information. See section 6.3 regarding caching issues.

To dynamically detect subnetwork-layer filtering in NBMA subnetworks (e.g., X.25 closed user group facility, or SMDS address screens), as well as to provide loop detection and diagnostic capabilities, NHRP optionally incorporates a "Route Record" in requests and replies (see Sections 5.7.4 and 5.7.5). The Route Record extensions contain the internetwork (and subnetwork layer) addresses of all intermediate NHSs between source and destination (in the forward direction) and between destination and source (in the reverse direction). When a source station is unable to communicate with the responder (e.g., an attempt to open an SVC fails), it may attempt to do so successively with other subnetwork layer addresses in the Route Record until it succeeds (if authentication policy permits such action). This approach can find a suitable egress point in the presence of subnetwork-layer filtering (which may be source/destination sensitive, for instance, without necessarily creating separate logical NBMA subnetworks) or subnetwork-layer congestion (especially in connection-oriented media).

NHRP messages, with the exception of Purge packets, are sent unreliably. NHRP requests should be retransmitted periodically until either a Reply or an Error packet is received.

3. Modes of Deployment

NHRP supports two deployment modes of operation: "server" and "fabric" modes. The two modes differ only in the way NHRP packets are propagated, which is driven by differences in configuration.

It is desirable that hosts attached directly to the NBMA subnetwork have no knowledge of whether NHRP is deployed in "server" or "fabric" modes, so that a change in deployment strategy can be done within a single administration, transparently to hosts. For this reason, host configuration is invariant between the two cases. Note that irrespective of which mode is deployed, NHRP clients must nominally be configured with the NBMA (and IP) address of at least one NHS. In practice, a host's default router should also be its NHS.

Server Mode

In "server" mode, the expectation is that a small number of NHSs will be fielded in an NBMA subnetwork. This may be appropriate in subnetworks containing routers that do not support NHRP, or

subnetworks that have large numbers of directly-attached hosts (and relatively few routers). Server mode assumes that NHRP is very loosely coupled with IP routing, and that the path taken by NHRP requests has little to do with the path taken by IP data packets routed to the desired destination. Note that in server mode the NHSs need not be routers, since they will not be required to forward transit data packets.

[Note: This is the likely scenario for initial deployment of NHRP. It is also likely that single and Multi-LIS configurations using either group-addressed ARP (in the case of SMDS) or ARP servers (in the case of ATM or SMDS) may already be in place.]

Server mode uses static configuration of NHS identity. The client station must be configured with the IP address of one or more NHSs, and there must be a path to that NHS (either directly, in which case the NHS's NBMA address must be known, or indirectly, through a router whose NBMA address is known). If there are multiple NHSs, they must be configured with each others' addresses, the identities of the destinations that each of them serves, and optionally a logical NBMA subnetwork identifier. (This static configuration requirement, which may involve authentication as well as addressing information, tends to limit such deployments to a very small number of NHSs.)

If the NBMA subnetwork offers a group addressing or multicast feature, the client (station) may be configured with a group address assigned to the group of next-hop servers. The client might then submit NHRP requests to the group address, eliciting a response from one or more NHSs, depending on the response strategy selected. Note that the constraints described in Section 2 regarding direct replies may apply.

The servers can also be deployed with the group or multicast address of their peers, and an NHS might use this as a means of forwarding NHRP requests it cannot satisfy to its peers. This might elicit a response (to the NHS) from one or more NHSs, depending on the response strategy. The NHS would then forward the NHRP reply to the NRHP request originator. The purpose of using group addressing or a similar multicast mechanism in this scenario would be to eliminate the need to preconfigure each NHS in a logical NBMA subnetwork with both the individual identities of other NHSs as well as the destinations they serve. It reduces the number of NHSs that might be traversed to process an NHRP request (in those configurations where NHSs either respond or forward via the multicast, only two NHSs would be traversed), and allows the NHS that serves the NHRP request originator to cache next hop information associated with the reply (again, within the

constraints described in Section 2).

The NHRP request packet's destination IP address is set by the source station to the first-hop NHS's IP address. If the addressed NHS does not serve the destination, the NHRP request is forwarded to the IP address of the NHS that serves the destination.

The responding NHS uses the source address from within the NHRP packet (not the source address of the IP packet) as the IP destination of the NHRP reply.

Note that, in many cases, NHSs deployed in Server Mode are unlikely to be able to resolve the next hop of destination that lies outside of the NBMA subnetwork, since doing so requires routing knowledge that is only provided by certain protocols (Link State routing protocols, for example); with many routing protocols, only the egress router itself knows that it is the egress router. The identity of the egress router may be provided by a server if such information is very static; in practical terms the egress router can only be guaranteed to be fixed if static routing is in use, or there is only one egress router. If the identity of egress routers cannot be determined, then the NHSs can only provide information about destinations directly attached to the NBMA subnetwork.

Fabric Mode

In "fabric" mode, it is expected that NHRP-capable routers are ubiquitous throughout the NBMA subnetwork, and that NHSs acquire knowledge about destinations other NHSs serve as a direct consequence of participating in intradomain and interdomain routing protocol exchange. In particular, the NHS serving a particular destination must lie along the routed path to that destination. In practice, this means that all egress routers must double as NHSs serving the destinations beyond them, and that hosts on the NBMA subnetwork are served by routers that double as NHSs.

Fabric mode leverages a routed infrastructure that "overlays" the NBMA subnetwork. The source station passes the NHRP request to the router which serves as the next hop toward the destination. Each router in turn forwards the NHRP request toward the destination. Eventually, the NHRP request arrives at a router that is acting as an NHS serving the destination (or the destination itself, if it is an NHRP-speaker), which generates the NHRP reply.

If the source station is a host, it sets the IP destination address of the NHRP request to the first-hop NHS/router (so that hosts needn't know the mode in which the subnetwork is running). If the

source station is a router, the destination IP address may be set either to the next-hop router or to the ultimate destination being resolved. Each NHS/router examines the NHRP request packet on its way toward the destination, optionally modifying it on the way (such as updating the Forward Record extension). The Router Alert option [5] is added by the first NHS in order to ensure that NHS/routers along the path process the packet, even though it may be addressed to the ultimate destination.

If an NHS/router receives an NHRP packet addressed to itself to which it cannot reply (because it does not serve the destination directly), it will forward the NHRP request with the destination IP address set to the ultimate destination (thus allowing invariant host behavior). Eventually, the NHRP packet will arrive at the NHS/router that serves the destination (which will return a positive NHRP reply) or it will arrive at a NHS/router that has no route to the destination (which will return a negative NHRP reply), or it may arrive at a NHS/router that cannot reach the NHS that serves the destination due to a loss of reachability among the NHSs (in which case the router will return a negative NHRP reply).

The procedural difference between server mode and fabric mode is reduced to deciding how to update the destination address in the IP packet carrying the NHRP request.

Note that addressing the NHRP request to the ultimate destination allows for subnetworks that do not have NHSs deployed in all routers; typically a very large NBMA subnetwork might only deploy NHSs in egress routers, and not in transit routers.

4. Configuration

Stations

To participate in NHRP, a station connected to an NBMA subnetwork should be configured with the IP and NBMA address(es) of its NHS(s) (alternatively, it should be configured with a means of acquiring them, i.e., the group address that members of a NHS group use for the purpose of address or next-hop resolution.) The NHS(s) may be physically located on the stations's default or peer routers, so their addresses may be obtained from the station's IP forwarding table. If the station is attached to several subnetworks (including logical NBMA subnetworks), the station should also be configured to receive routing information from its NHS(s) and peer routers so that it can determine which IP networks are reachable through which subnetworks.

Next Hop Servers

An NHS is configured with its own identity, a set of IP address prefixes that correspond to the IP addresses of the stations it serves, a logical NBMA subnetwork identifier (see Section 5.7.2), and in the case of "server" mode, the identities of other NHSs in the same logical NBMA subnetwork. If a served station is attached to several subnetworks, the NHS may also need to be configured to advertise routing information to such stations.

If an NHS acts as an egress router for stations connected to other subnetworks than the NBMA subnetwork, the NHS must, in addition to the above, be configured to exchange routing information between the NBMA subnetwork and these other subnetworks.

In all cases, routing information is exchanged using conventional intra-domain and/or inter-domain routing protocols.

The NBMA addresses of the stations served by the NHS may be learned via NHRP Register packets or manual configuration.

5. Packet Formats

This section describes the format of NHRP packets.

An NHRP packet consists of a Fixed Part, a Mandatory Part, and an Extensions Part. The Fixed Part is common to all NHRP packet types. The Mandatory Part MUST be present, but varies depending on packet type. The Extensions Part also varies depending on packet type, and need not be present.

The length of the Fixed Part is fixed at 8 octets. The length of the Mandatory Part is carried in the Fixed Part. The length of the Extensions Part is implied by the total packet length (Internet datagram total length minus IP header length minus NHRP fixed part length minus NHRP mandatory part length).

NHRP packets are carried in IP packets as protocol type 54 (decimal). NHSs may increase the size of an NHRP packet as a result of extension processing. IP datagrams containing NHRP packets MUST have the Don't Fragment bit set.

Fields marked "unused" MUST be set to zero on transmission, and ignored on receipt.

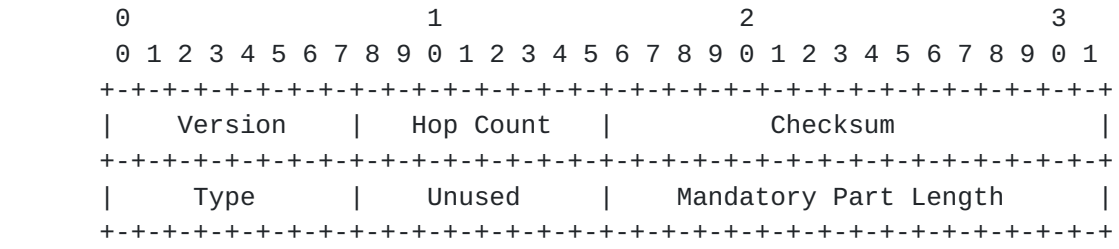
Most packet types have both internetwork layer protocol-independent fields and protocol-specific fields. The protocol-independent fields

always come first in the packet, and the Protocol ID field qualifies the format of the protocol-specific fields. The protocol-specific fields defined in this document are for IPv4 only; formats of protocol-specific fields for other protocols are for further study.

The protocol ID field in general will contain the Ethertype value for the protocol (see [6]). For protocols that do not have an assigned Ethertype, this field will in general contain the Network Layer Protocol Identifier (NLPID, [7]) value for the protocol (this is guaranteed to not cause collisions since the NLPID cannot be greater than 255 decimal, and the Ethertype cannot be less than 1500 decimal).

5.1 NHRP Fixed Header

The NHRP Fixed Header is present in all NHRP packets. It contains the basic information needed to parse the rest of the packet.



Version
The NHRP version number. Currently this value is 1.

Hop Count
The Hop count indicates the maximum number of NHSs that an NHRP packet is allowed to traverse before being discarded.

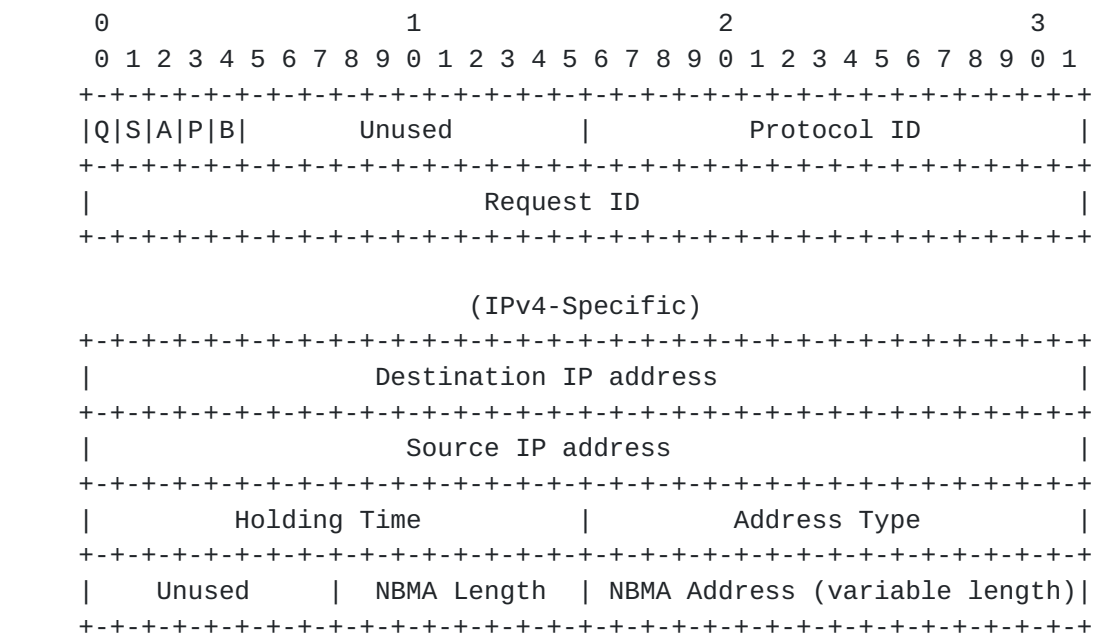
Checksum
The standard IP checksum over the entire NHRP packet (starting with the fixed header). If only the hop count field is changed, the checksum is adjusted without full recomputation. The checksum is completely recomputed when other header fields are changed.

Type
The NHRP packet type: Request, Response, Register, Purge, or Error Indication (see below).

Mandatory Part Length
The length in octets of the Mandatory Part. This length does not include the Fixed Header.

5.2 NHRP Request

The NHRP Request packet has a Type code of 1. The Mandatory Part has the following format:



Q
Set if the Requestor is a router; clear if the requestor is a host.

S
Unused (zero on transmit)

A
A response to an NHRP request may contain cached information. If an authoritative answer is desired, then this bit ("Authoritative") should be set. If non-authoritative (cached) information is acceptable, this bit should be clear.

P
Unused (zero on transmit)

B
Unused (zero on transmit)

Protocol ID
Specifies the internetwork layer protocol for which we are obtaining routing information. This value also qualifies the structure of the remainder of the Mandatory Part. For IPv4, the Protocol ID is hexadecimal 800 (decimal 2048). Protocol ID values

for other internetwork layer protocols are for future study.

Request ID
A value which, when coupled with the address of the source, provides a unique identifier for the information contained in a Request and its associated Reply, and any subsequent Purge. This value can be used by the source to aid in matching requests with replies. This value could also be sent across a virtual circuit (in SVC environments) to aid in matching NHRP transactions with virtual circuits (this use is for further study).

The value is taken from a 32 bit counter that is incremented each time a new NHRP request is transmitted. The same value MUST be used when sending another request for the same destination when a previous request is still active or pending, i.e., when retransmitting a request because a reply was not received, or when refreshing an existing entry to avoid holding timer expiration. A new value MUST be used when sending a request when no cache entry is present, or a previous cache entry was deleted for any reason.

Destination and Source IP Addresses
Respectively, these are the IP addresses of the station for which the NBMA next hop is desired, and the NHRP request initiator.

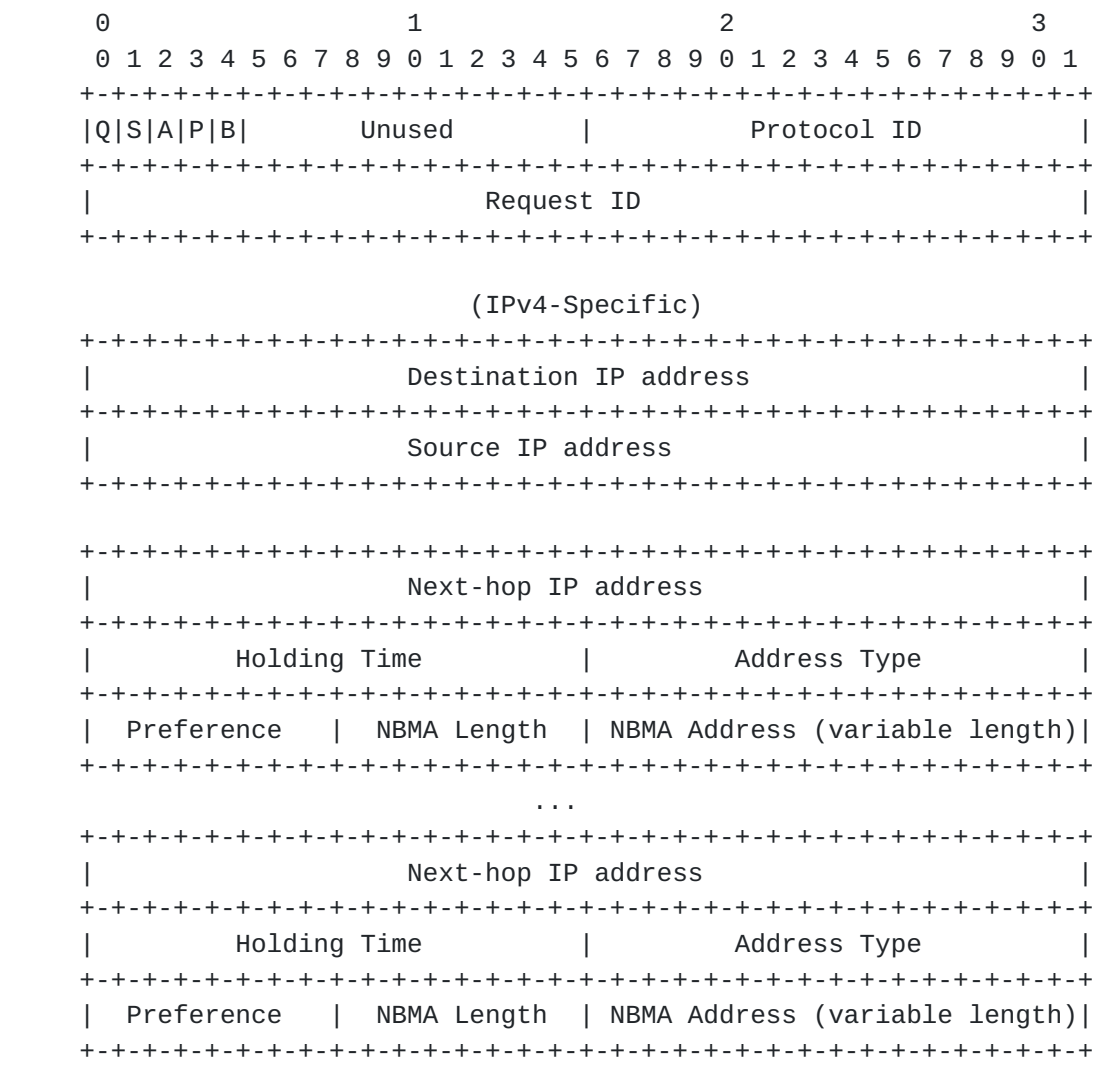
Source Holding Time, Address Type, NBMA Length, and NBMA Address
The Holding Time field specifies the number of seconds for which the source NBMA information is considered to be valid. Cached information SHALL be discarded when the holding time expires.

The Address Type field specifies the type of NBMA address (qualifying the NBMA address). Possible address types are listed in [6].

The NBMA length field is the length of the NBMA address of the source station in bits. The NBMA address field itself is zero-filled to the nearest 32-bit boundary.

5.3 NHRP Reply

The NHRP Reply packet has a type code of 2. The Mandatory Part has the following format:



Q
Copied from the NHRP Request. Set if the Requestor is a router; clear if the requestor is a host.

S
Set if the next hop identified in the reply is a router; clear if the next hop is a host.

A
Set if the reply is authoritative; clear if the reply is non-authoritative.

P
Set if the reply is positive; clear if the reply is negative.

B
Set if the association between the destination and the next hop information is guaranteed to be stable for the lifetime of the information (the holding time). This is the case if the Next-hop IP address identifies the destination (though it may be different in value than the Destination address if the destination system has multiple addresses) or if the destination is not connected directly to the NBMA subnetwork but the egress router to that destination is guaranteed to be stable (such as when the destination is immediately adjacent to the egress router through a non-NBMA interface). This information affects cacheing strategies (see section 6.3).

An NHS is not allowed to reply to an NHRP request for authoritative information with cached information, but may do so for an NHRP Request which indicates a request for non-authoritative information. An NHS may reply to an NHRP request for non-authoritative information with authoritative information.

Protocol ID
Specifies the internetwork layer protocol for which we are obtaining routing information. This value also qualifies the structure of the remainder of the Mandatory Part. For IPv4, the Protocol ID is hexadecimal 800 (decimal 2048). Protocol ID values for other internetwork layer protocols are for future study.

Request ID
Copied from the NHRP Request.

Destination IP Address
The address of the target station (copied from the corresponding NHRP Request).

Source IP Address
The address of the initiator of the request (copied from the corresponding NHRP Request).

Next-hop entry
A Next-hop entry consists of the following fields: a 32-bit Next-hop IP Address, a 16-bit Holding Time, an 8-bit Preference, an 8-bit Address Type, an 8-bit NBMA Length, and an NBMA Address whose length is the value of the NBMA length field.

The Next-hop IP Address specifies the IP address of the next hop. This will be the address of the destination host if it is directly attached to the NBMA subnetwork, or the egress router if it is not directly attached.

The Holding Time field specifies the number of seconds for which the associated Next-hop entry information is considered to be valid. Cached information SHALL be discarded when the holding time expires. (Holding time is to be specified for both positive and negative replies).

The Address Type field specifies the type of NBMA address (qualifying the NBMA address). Possible address types are listed in [6].

The Preference field specifies the preference of the Next-hop entry, relative to other Next-hop entries in this NHRP Reply packet. Higher values indicate more preferable Next-hop entries. Action taken when multiple next-hop entries have the highest preference value is a local matter.

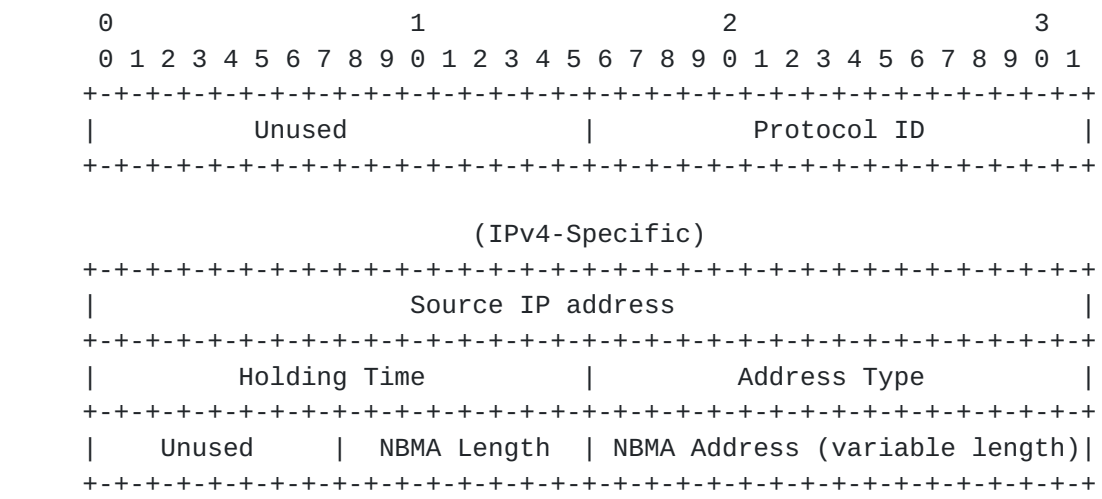
The NBMA length field specifies the length of the NBMA address of the destination station in bits. The NBMA address field itself is zero-filled to the nearest 32-bit boundary. For negative replies, the Holding Time field is relevant; however, the preference, Address Type, and NBMA length fields MUST be zero, and the NBMA Address SHALL NOT be present.

There may be multiple Next-hop entries returned in the reply (as implied by the Mandatory Part Length). The preference values are used to select the preferred entry. The same next-hop IP address may be associated with multiple NBMA addresses. Load-splitting may be performed over the addresses, given equal preference values, and the alternative addresses may be used in case of connectivity failure in the NBMA subnetwork (such as a failed call attempt in connection-oriented NBMA subnetworks).

If extensions were present in the NHRP Request packet, all of these extensions MUST be present in the NHRP Reply. No additional extensions may be added to the reply that were not present in the request.

5.4 NHRP Register

The NHRP Register packet is sent from a station to an NHS to notify the NHS of the station's NBMA address. It has a Type code of 3. The Mandatory Part has the following format:



Protocol ID
Specifies the internetwork layer protocol for which we are obtaining routing information. This value also qualifies the structure of the remainder of the Mandatory Part. For IPv4, the Protocol ID is hexadecimal 800 (decimal 2048). Protocol ID values for other internetwork layer protocols are for future study.

Source IP Address
The IP address of the station wishing to register its NBMA address with an NHS.

Source Holding Time, Address Type, NBMA Length, and NBMA Address
The Holding Time field specifies the number of seconds for which the source NBMA information is considered to be valid. Cached information SHALL be discarded when the holding time expires.

The Address Type field specifies the type of NBMA address (qualifying the NBMA address). Possible address types are listed in [6].

The NBMA length field is the length of the NBMA address of the source station in bits. The NBMA address itself is zero-filled to the nearest 32-bit boundary.

This packet is used to register a station's IP and NBMA addresses with its configured NHS. This allows static configuration information to be reduced; the NHSs need not be configured with the identities of all of the stations that they serve.

It is possible that a misconfigured station will attempt to register with the wrong NHS (i.e., one that cannot serve it due to policy constraints or routing state). If this is the case, the NHS MUST

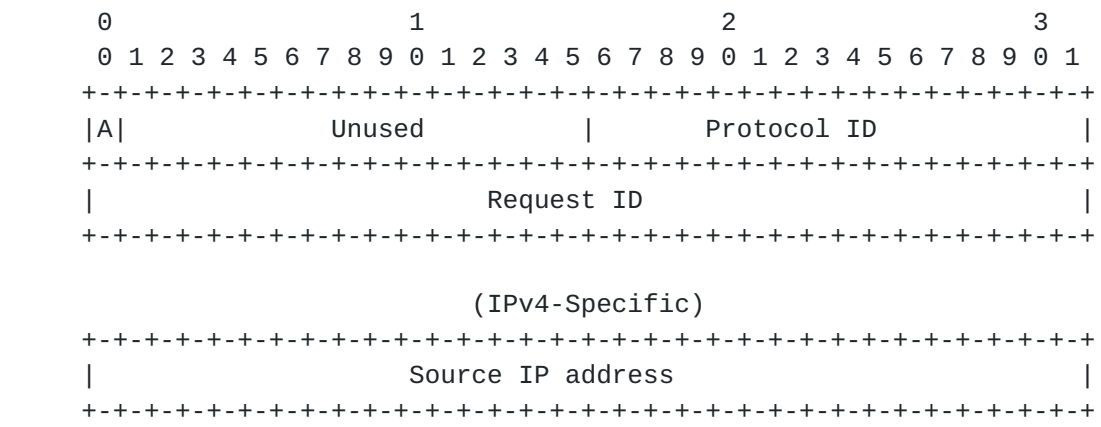
reply with an Error Indication of type Can't Serve This Address.

If an NHS cannot serve a station due to a lack of resources, the NHS MUST reply with an Error Indication of type Registration Overflow.

In order to keep the registration entry from being discarded, the station MUST resend the Register packet often enough to refresh the registration, even in the face of occasional packet loss. It is recommended that the Registration packet be sent at an interval equal to one-third of the Holding Time specified therein.

5.5 NHRP Purge

The NHRP Purge packet has a type code of 4. The Mandatory Part has the following format:



A
Clear if this is a purge request, set if this is an acknowledgement.

Protocol ID
Specifies the internetwork layer protocol for which we are obtaining routing information. This value also qualifies the structure of the remainder of the Mandatory Part. For IPv4, the Protocol ID is hexadecimal 800 (decimal 2048). Protocol ID values for other internetwork layer protocols are for future study.

Request ID
Copied from the corresponding NHRP Request. This is used by the station receiving the purge to identify which cache entry to purge, and by the NHS receiving the acknowledgement to match the

acknowledgement with the Purge request.

Source IP Address
The address of the initiator of the request (copied from the corresponding NHRP Request). Used by the NHS receiving the acknowledgement to match the acknowledgement with the Purge request.

An NHRP Purge request packet is sent from an NHS to a station to cause it to delete previously cached information. This is done when the information may be no longer valid (typically when the NHS has previously provided next hop information for a destination that is not directly connected to the NBMA subnetwork, and the egress point to that destination may have changed).

The IP destination address of the packet containing the Purge request is set to the Source IP address from the original Request packet.

The NHS sending the NHRP Purge request MUST periodically retransmit the request until it is acknowledged, or until the holding time of the information being purged has expired. Retransmission strategies are for further investigation.

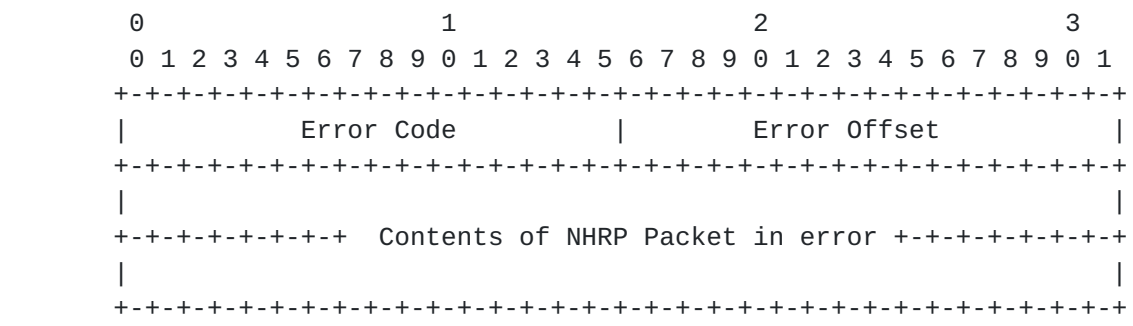
When a station receives an NHRP Purge request, it MUST discard any previous cached information that matches the Request ID. It MUST then acknowledge the Purge request by setting the Acknowledgement (A) bit and returning the Purge request to the sender. The IP destination address of the Purge acknowledgement MUST be set to the IP source address of the Purge request.

An acknowledgement MUST be returned for the Purge request even if the station does not have a cache entry with a matching Request ID.

If the station wishes to reestablish communication with the destination shortly after receiving a Purge request, it should make an authoritative request in order to avoid any stale cache entries that might be present in intermediate NHSs. (See section 6.3.2.) It is recommended that authoritative requests be made for the duration of the holding time of the old information.

5.6 NHRP Error Indication

The NHRP Error Indication is used to convey error indications to the initiator of an NHRP Request packet. It has a type code of 5. The Mandatory Part has the following format:



Error Code
An error code indicating the type of error detected, chosen from the following list:

1 - Unrecognized Extension
2 - Subnetwork ID Mismatch
3 - NHRP Loop Detected
4 - Can't Serve This Address
5 - Registration Overflow
6 - Server Unreachable
7 - Protocol Error
8 - NHRP fragmentation failure

Error Offset
The offset in octets into the original NHRP packet, starting at the NHRP Fixed Header, at which the error was detected.

The destination IP address of an NHRP Error Indication SHALL be set to the IP address of the initiator of the original NHRP Request (as extracted from the NHRP Request or NHRP Reply).

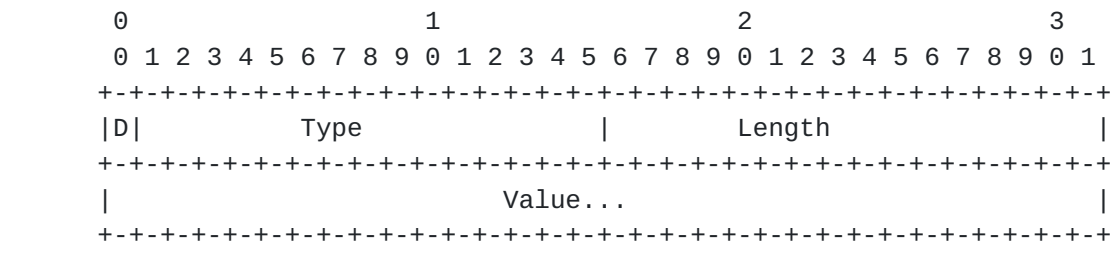
An Error Indication packet SHALL NEVER be generated in response to another Error Indication packet. When an Error Indication packet is generated, the offending NHRP packet SHALL be discarded. In no case should more than one Error Indication packet be generated for a single NHRP packet.

5.7 Extensions Part

The Extensions Part, if present, carries one or more extensions in {Type, Length, Value} triplets. Extensions are only present in a Reply if they were present in the corresponding Request; therefore, minimal NHRP station implementations that do not act as an NHS and do not transmit extensions need not be able to receive them. An implementation that is incapable of processing extensions SHALL return an Error Indication of type Unrecognized Extension when it receives an NHRP packet containing extensions.

Extensions are typically protocol-specific, as noted.

Extensions have the following format:



D
"Discretionary." If set, and the NHS does not recognize the type code, the extension may safely be ignored. If clear, and the NHS does not recognize the type code, the NHRP request is considered in error. (See below for details.)

Type
The extension type code (see below). The extension type is not qualified by the Discretionary bit, but is orthogonal to it.

Length
The length in octets of the value (not including the Type and Length fields; a null extension will have only an extension header and a length of zero).

Each extension is padded with zero octets to a 32 bit boundary. This padding is not included in the Length field.

Extensions may occur in any order, but any particular extension type may occur only once in an NHRP packet.

The Discretionary bit provides for a means to add to the extension set. If the bit is clear, the NHRP request cannot be satisfied unless the extension is processed, so the responder MUST return an Error Indication of type Unrecognized Extension. If the bit is set, the extension can be safely ignored, though unrecognized extensions so ignored that were received in an NHRP Request packet MUST be returned unchanged in the corresponding NHRP Reply.

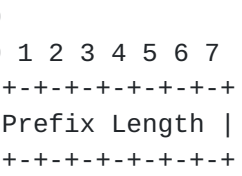
If a transit NHS (one which is not going to generate a reply) detects an unrecognized extension, it SHALL ignore the extension. If the Discretionary bit is clear, the transit NHS MUST NOT cache the information (in the case of a reply) and MUST NOT identify itself as an egress router (in the Forward Record or Reverse Record extensions). Effectively, this means that a transit NHS that encounters an extension that it cannot process and determines that

the Discretionary bit is clear MUST NOT participate in any way in the protocol exchange, other than acting as a forwarding agent for the request.

5.7.1 Destination Prefix Extension (IPv4-Specific)

Discretionary = 1
Type = 1
Length = 1

This extension is used to indicate that the information carried in an NHRP Reply pertains to an equivalence class of destinations rather than just the destination IP address specified in the request. All addresses that match the IP address prefix defined by the prefix length are part of the equivalence class.



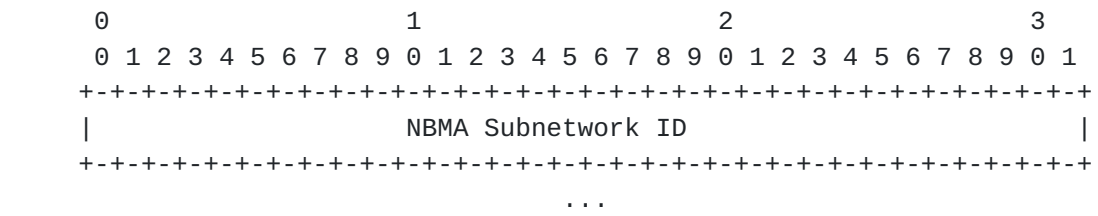
If an initiator would like to receive this equivalence information, it SHALL add this extension to the NHRP Request with a value of 32. The responder SHALL copy the extension to the NHRP Reply and modify the prefix length appropriately.

5.7.2 NBMA Subnetwork ID Extension (Protocol-Independent)

Discretionary = 0
Type = 2
Length = variable

This extension is used to carry one or more identifiers for the NBMA subnetwork. This can be used as a validity check to ensure that the request does not leave a particular NBMA subnetwork. The extension is placed in an NHRP Request packet by the initiator with an ID value of zero; the first NHS fills in the field with the identifier(s) for the NBMA subnetwork.

Multiple NBMA Subnetwork IDs may be used as a transition mechanism while NBMA Subnetworks are being split or merged.



Each identifier consists of a 32 bit globally unique value assigned to the NBMA subnetwork. This value should be chosen from the IP address space administered by the operators of the NBMA subnetwork. This value is used for identification only, not for routing or any other purpose.

Each NHS processing an NHRP Request SHALL verify these values. If none of the values matches the NHS's NBMA Subnetwork ID, the NHS SHALL return an Error Indication of type "Subnetwork ID Mismatch" and discard the NHRP Request.

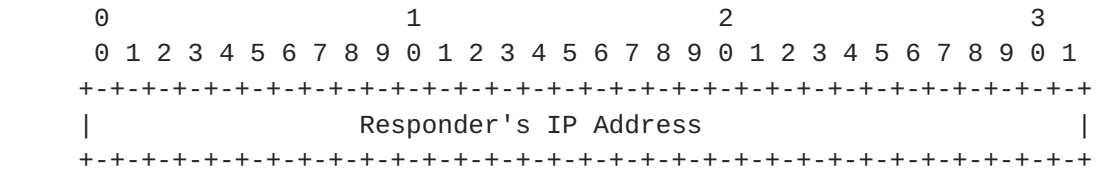
When an NHS is building an NHRP Reply and the NBMA Subnetwork ID extension is present in the NHRP Request, the NBMA Subnetwork ID extension SHALL be copied from the Request to the Reply, including all values carried therein.

Each NHS processing an NHRP Reply SHALL verify the values carried in the NBMA Subnetwork ID extension, if present. If none of the values matches the NHSs NBMA Subnetwork ID, the NHS SHALL return an Error Indication of type "Subnetwork ID Mismatch" and discard the NHRP Reply.

5.7.3 Responder Address Extension (IPv4-Specific)

Discretionary = 0
Type = 3
Length = 4

This extension is used to determine the IP address of the NHRP Responder, that is, the entity that generates the NHRP Reply packet. The intent is to identify the entity responding to the request, which may be different (in the case of cached replies) than the system identified in the Next-hop field of the reply, and to aid in detecting loops in the NHRP forwarding path.



If a requestor desires this information, it SHALL include this extension, with a value of zero, in the NHRP Request packet.

If an NHS is generating an NHRP Reply packet in response to a request containing this extension, it SHALL include this extension, containing its IP address, in the NHRP Reply. If an NHS has more than one IP address, it SHALL use the same IP address consistently in all of the Responder Address, Forward NHS Record, and Reverse NHS Record extensions. The choice of which of several IP addresses to include in this extension is a local matter.

If an NHRP Reply packet being forwarded by an NHS contains an IP address of that NHS in the Responder Address Extension, the NHS SHALL generate an Error Indication of type "NHRP Loop Detected" and discard the Reply.

If an NHRP Reply packet is being returned by an intermediate NHS based on cached data, it SHALL place its own address in this extension (differentiating it from the address in the Next-hop field).

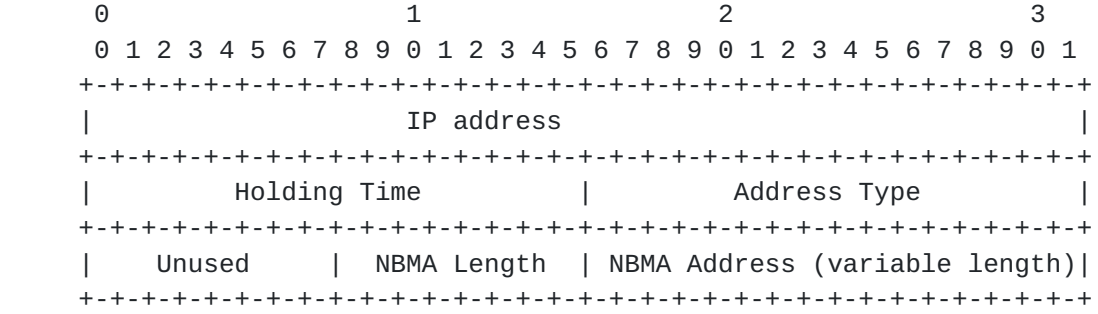
5.7.4 NHRP Forward NHS Record Extension (IPv4-Specific)

Discretionary = 0
Type = 4
Length = variable

The NHRP forward NHS record is a list of NHSs through which an NHRP request traverses. Each NHS SHALL append a Next-hop element containing its IP address to this extension.

In addition, NHSs that are willing to act as egress routers for packets from the source to the destination SHALL include information about their NBMA Address.

Each Next-hop element is formatted as follows:



IP address
The IP address of the NHS.

Holding Time
The number of seconds for which this information is valid. If a station chooses to use this information as a next-hop entry, it may not be used once the holding timer expires.

Address Type, NBMA Length, and NBMA Address
The Address Type field specifies the type of NBMA address (qualifying the NBMA address). Possible address types are listed in [6].

The NBMA length field is the length of the NBMA address of the destination station in bits. The NBMA address itself is zero-filled to the nearest 32-bit boundary.

NHSs that are not egress routers SHALL specify an NBMA Length of zero and SHALL NOT include an NBMA Address.

If a requestor wishes to obtain this information, it SHALL include this extension with a length of zero.

Each NHS SHALL append an appropriate Next-hop element to this extension when processing an NHRP Request. The extension length field and NHRP checksum SHALL be adjusted as necessary.

The last-hop NHS (the one that will be generating the NHRP Reply) SHALL NOT update this extension (since this information will be in the reply).

If an NHS has more than one IP address, it SHALL use the same IP address consistently in all of the Responder Address, Forward NHS Record, and Reverse NHS Record extensions. The choice of which of several IP addresses to include in this extension is a local matter.

If an NHRP Request packet being forwarded by an NHS contains the IP address of that NHS in the Forward NHS Record Extension, the NHS SHALL generate an Error Indication of type "NHRP Loop Detected" and discard the Request.

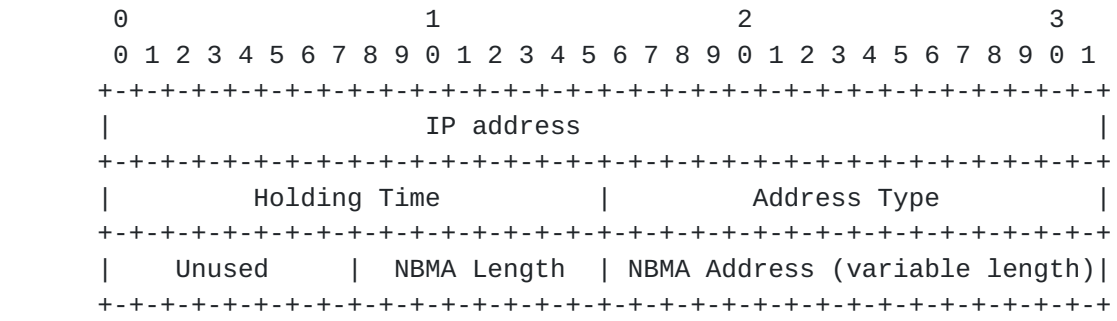
5.7.5 NHRP Reverse NHS Record Extension (IPv4-Specific)

Discretionary = 0
Type = 5
Length = variable

The NHRP reverse NHS record is a list of NHSs through which an NHRP reply traverses. Each NHS SHALL append a Next-hop element containing its IP address to this extension.

In addition, NHSs that are willing to act as egress routers for packets from the source to the destination SHALL include information about their NBMA Address.

Each Next-hop element is formatted as follows:



IP address
The IP address of the NHS.

Holding Time
The number of seconds for which this information is valid. If a station chooses to use this information as a next-hop entry, it may not be used once the holding timer expires.

Address Type, NBMA Length, and NBMA Address
The Address Type field specifies the type of NBMA address (qualifying the NBMA address). Possible address types are listed in [6].

The NBMA length field is the length of the NBMA address of the destination station in bits. The NBMA address itself is zero-filled to the nearest 32-bit boundary.

NHSs that are not egress routers SHALL specify an NBMA Length of zero and SHALL NOT include an NBMA Address.

If a requestor wishes to obtain this information, it SHALL include this extension with a length of zero.

Each NHS SHALL append an appropriate Next-hop element to this extension when processing an NHRP Reply. The extension length field and NHRP checksum SHALL be adjusted as necessary.

The NHS generating the NHRP Reply SHALL NOT update this extension.

If an NHS has more than one IP address, it SHALL use the same IP address consistently in all of the Responder Address, Forward NHS Record, and Reverse NHS Record extensions. The choice of which of several IP addresses to include in this extension is a local matter.

If an NHRP Reply packet being forwarded by an NHS contains the IP address of that NHS in the Reverse NHS Record Extension, the NHS SHALL generate an Error Indication of type "NHRP Loop Detected" and discard the Reply.

Note that this information may be cached at intermediate NHSs; if so, the cached value SHALL be used when generating a reply. Note that the Responder Address extension may be used to disambiguate the set of NHSs that actually processed the reply.

5.7.6 NHRP QoS Extension

Discretionary = 1
Type = 6
Length = variable

The NHRP QoS Extension is carried in NHRP Request packets to indicate the desired QoS of the path to the indicated destination. This information may be used to help select the appropriate NBMA next hop.

It may also be carried in NHRP Register packets to indicate the QoS to which the registration applies.

The syntax and semantics of this extension are TBD; alignment with resource reservation may be useful.

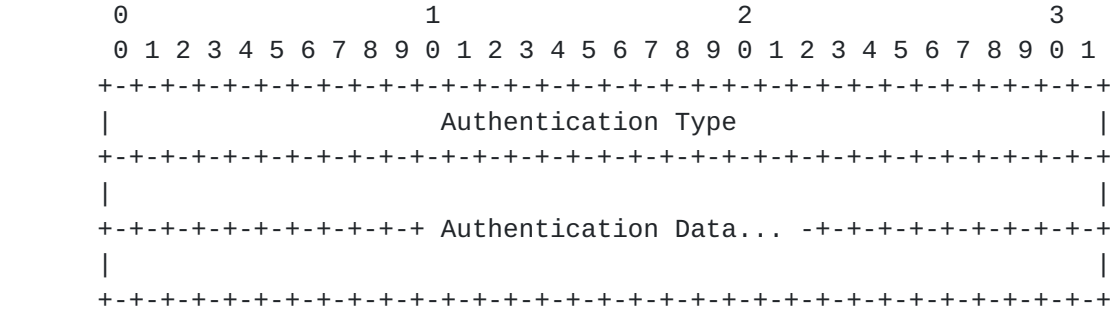
5.7.7 NHRP Authentication Extension

Discretionary = 0
Type = 7
Length = variable

The NHRP Authentication Extension is carried in NHRP packets to convey authentication information between NHRP speakers. The Authentication Extension may be included in any NHRP packet type.

Authentication is done pairwise on an NHRP hop-by-hop basis; the authentication extension is regenerated on each hop. If a received packet fails the authentication test, the NHS SHALL generate an Error Indication of type "Authentication Failure" and discard the packet. In no case SHALL an Error Indication packet be generated on the

receipt of an Error Indication packet, however. Note that one possible authentication failure is the lack of an Authentication Extension; the presence or absence of the Authentication Extension is a local matter.



The Authentication Type field identifies the authentication method in use. Currently assigned values are:

1 - Cleartext Password
2 - Keyed MD5

All other values are reserved.

The Authentication Data field contains the type-specific authentication information.

In the case of Cleartext Password Authentication, the Authentication Data consists of a variable length password.

In the case of Keyed MD5 Authentication, the Authentication Data contains the 16 byte MD5 digest of the entire NHRP packet, including the IP header, with the authentication key appended to the end of the packet. The authentication key is not transmitted with the packet.

Distribution of authentication keys is outside the scope of this document.

5.7.8 NHRP Vendor-Private Extension

Discretionary = 1
Type = 8
Length = variable

The NHRP Vendor-Private Extension is carried in NHRP packets to convey vendor-private information or NHRP extensions between NHRP speakers. This extension may be used at any time; if the receiver

does not handle this extension, or does not match the vendor ID in the extension, then the extension may be completely ignored by the receiver. The first 24 bits of the extension's payload (following the length field) contains the 802 vendor ID as assigned by the IEEE [6]. The remaining octets in the payload are vendor-dependent.

6. Protocol Operation

In this section, we discuss certain operational considerations of NHRP.

6.1 Router-to-Router Operation

In practice, the initiating and responding stations may be either hosts or routers. However, there is a possibility under certain conditions that a stable routing loop may occur if NHRP is used between two routers. In particular, attempting to establish an NHRP path across a boundary where information used in route selection is lost may result in a routing loop. Such situations include the loss of BGP path vector information, the interworking of multiple routing protocols with dissimilar metrics (e.g, RIP and OSPF), etc. In such circumstances, NHRP should not be used. This situation can be avoided if there are no "back door" paths between the entry and egress router outside of the NBMA subnetwork. Protocol mechanisms to relax these restrictions are under investigation.

In general it is preferable to use mechanisms, if they exist, in routing protocols to resolve the egress point when the destination lies outside of the NBMA subnetwork, since such mechanisms will be more tightly coupled to the state of the routing system and will probably be less likely to create loops.

6.2 Handling of IP Destination Address Field

NHRP packets are self-contained in terms of the IP addressing information needed for protocol operation--the IP source and destination addresses in the encapsulating IP header are not used. However, the setting of the IP destination address field does impact how NHRP requests are forwarded.

There are essentially three choices in how to set the destination IP address field at any particular point in the forwarding of an NHRP request: the ultimate destination being resolved, the next-hop IP router on the path to the destination, and the next-hop NHS (which might not be adjacent to the NHS forming the packet header).

The first case, addressing the packet to the destination being resolved (in the hopes that an NHS lies along the path) is desirable for at least two reasons. It simplifies configuration (since the identity of the next NHS need not be known explicitly), and it simplifies deployment (since the packet will pass silently through routers that are not NHSs). However, it assumes that the serving NHS lies along the path to the destination, and it requires NHSs along the path to examine the packet even though it is not addressed to them.

The second case, addressing the packet to the next-hop router, is similar to the first in that it follows the path to the destination, thus reducing configuration complexity. It furthermore only requires NHSs to process the packet if they are directly addressed. It too assumes that the responding NHS is on the path to the destination. However, it requires that all routers along the path are also NHSs.

The third case, addressing the packet to the next-hop NHS, allows the NHSs to be independent of routing, and requires only addressed NHSs to examine the packet. However, there is no reasonable way, other than manual configuration, to determine the identity of the next hop NHS if it is not also the next hop IP router (making it option two).

In order to balance all of these issues, the following rules SHALL be used when constructing IP packets to carry NHRP requests.

Stations

Stations SHALL address NHRP packets to the NHS by which they are served, regardless of whether NHRP has been deployed in Server mode or Fabric mode.

NHSs

If an NHS receives an NHRP packet in which the IP destination address does not match any of its own IP addresses, it SHALL process the NHRP packet as appropriate, and if it MUST forward the NHRP packet to another NHS, SHALL transmit the packet with the same IP destination address with which it was received.

If an NHS receives an NHRP packet in which the IP destination address matches one of its own IP addresses, it SHALL process the NHRP packet as appropriate, and if it MUST forward the NHRP packet to another NHS, SHALL set the destination IP address in one of the following ways:

If there is a configured next-hop NHS for the destination being

resolved (Server mode), it SHALL transmit the packet with the IP destination address set to the next-hop NHS.

If there is no configured next-hop NHS (Fabric Mode), it SHALL transmit the packet with the IP destination address set to the address of the destination being resolved, and SHALL include the Router Alert option [5] so that intermediate NHS/routers can examine the NHRP packet.

6.3 Cache Management Issues

The management of NHRP caches in the source station, the NHS serving the destination, and any intermediate NHSs is dependent on a number of factors.

6.3.1 Cacheing Requirements

Source Stations

Source stations must of course cache all received replies that they are actively using. They also must cache "incomplete" entries, i.e., those for which a request has been sent but which a reply has not been received. This is necessary in order to preserve the Request ID for retries, and provides the state necessary to avoid triggering requests for every data packet sent to the destination.

Source stations MUST purge expired information from their caches. Source stations MUST purge the appropriate cached information upon receipt of an NHRP Purge request packet.

Source stations that are also NHSs may return cached information learned in response to its own NHRP Request packets in reply to requests it receives, within the rules for Transit NHSs below.

Serving NHSs

The NHS serving the destination (the one which responds authoritatively to NHRP requests) MUST cache information about all requests to which it has responded if the information in the reply has the possibility of changing during its lifetime (so that an NHRP Purge request packet can be sent). The NBMA information provided by the source station in the NHRP Request may be cached for the duration of its holding time. This information is considered to be stable, since it identifies a station directly attached to the NBMA subnetwork.

Transit NHSs

A Transit NHS (lying along the NHRP path between the source station and the responding NHS) may cache information contained in NHRP Request packets that it forwards. A Transit NHS may cache information contained in NHRP Reply packets that it forwards only if that reply has the Stable (B) bit set. It MUST discard any cached information whose holding time has expired. It may return cached information in response to non-authoritative requests only.

6.3.2 Dynamics of Cached Information

NBMA-Connected Destinations

NHRP's most basic function is that of simple NBMA address resolution of stations directly attached to the NBMA subnetwork. These mappings are typically very static, and appropriately chosen holding times will minimize problems in the event that the NBMA address of a station must be changed. Stale information will cause a loss of connectivity, which may be used to trigger an authoritative NHRP request and bypass the old data. In the worst case, connectivity will fail until the cache entry times out.

This applies equally to information marked in replies as being "stable" (via the "B" bit).

This also applies equally well to source stations that are routers as well as those which are hosts.

Note that the information carried in the NHRP Request packet is always considered "stable" because it represents a station that is directly connected to the NBMA subnetwork.

Destinations Off of the NBMA Subnetwork

If the source of a request is a host and the destination is not directly attached to the NBMA subnetwork and is not considered to be "stable," the destination mapping may be very dynamic (except in the case of a subnetwork where each destination is only singly homed to the NBMA subnetwork). As such the cached information may very likely become stale. The consequence of stale information in this case will be a suboptimal path (unless the internetwork has partitioned or some other routing failure has occurred).

If the egress router/NHS detects a routing change toward the destination, it MUST send an NHRP Purge packet to the source, which

will usually cause the source to issue a authoritative request and find the new egress point. If the egress router for some reason sees no change in routing toward the destination, then it should still be a viable, if suboptimal, hop toward the destination. The consequence of the source making a non-authoritative request after a Purge, in the presence of stale cache entries (not removed by a Purge), is also suboptimal routing.

6.4 Use of the Destination Prefix Extension

A certain amount of care needs to be taken when using the Destination Prefix Extension, in particular with regard to the prefix length advertised (and thus the size of the equivalence class specified by it). Assuming that the routers on the NBMA subnetwork are exchanging routing information, it should not be possible for an NHS to create a black hole by advertising too large of a set of destinations, but suboptimal routing can result. For example, it should not be assumed that the proper prefix to advertise is the one provided by the routing system (especially if the prefix is determined from the default route).

The approach used to determine the prefix width is likely to vary based on the particulars of the situation. Information could be gleaned from local topology, routing protocols, and other sources.

In general, the width of the prefix should be handled conservatively (erring toward a longer prefix).

If multiple cache entries match the desired destination address (due to overlapping prefixes), the longest prefix MUST be used.

7. Security Considerations

As in any routing protocol, there are a number of potential security attacks possible, particularly denial-of-service attacks. The use of authentication on all packets is recommended to avoid such attacks.

The authentication schemes described in this document are intended to allow the receiver of a packet to validate the identity of the sender; they do not provide privacy or protection against replay attacks.

Detailed security analysis of this protocol is for further study.

8. Discussion

The result of an NHRP request depends on how routing is configured among the NHSs of an NBMA subnetwork. If the destination station is directly connected to the NBMA subnetwork and the the routed path to it lies entirely within the NBMA subnetwork, the NHRP replies always return the NBMA address of the destination station itself rather than the NBMA address of some egress router. On the other hand, if the routed path exits the NBMA subnetwork, NHRP will be unable to resolve the NBMA address of the destination, but rather will return the address of the egress router. For destinations outside the NBMA subnetwork, egress routers and routers in the other subnetworks should exchange routing information so that the optimal egress router may be found.

When the NBMA next hop toward a destination is not the destination station itself, the optimal NBMA next hop may change dynamically. This can happen, for instance, when an egress router nearer to the destination becomes available. This change can be detected in a number of ways. First of all, the source station will need to periodically reissue the NHRP Request at a minimum just prior to the expiration of the holding timer. Alternatively, the source can be configured to receive routing information from the routing system. When it detects an improvement in the route to the destination, the source can reissue the NHRP request to obtain the current optimal NBMA next hop. Source stations that are routers may choose to establish a routing association with the egress router, allowing the egress router to explicitly inform the source about changes in routing (and providing additional routing information, authentication, etc.) Such strategies will be discussed in a separate document.

In addition to NHSs, an NBMA station could also be associated with one or more regular routers that could act as "connectionless servers" for the station. The station could then choose to resolve the NBMA next hop or just send the IP packets to one of its connectionless servers. The latter option may be desirable if communication with the destination is short-lived and/or doesn't require much network resources. The connectionless servers could, of course, be physically integrated in the NHSs by augmenting them with IP switching functionality.

References

[1] NBMA Address Resolution Protocol (NARP), Juha Heinanen and Ramesh Govindan, draft-ietf-rolc-nbma-arp-00.txt.

[2] Address Resolution Protocol, David C. Plummer, RFC 826.

[3] Classical IP and ARP over ATM, Mark Laubach, RFC 1577.

[4] Transmission of IP datagrams over the SMDS service, J. Lawrence and D. Piscitello, RFC 1209.

[5] IP Router Alert Option, Dave Katz, draft-katz-router-alert-00.txt.

[6] Assigned Numbers, J. Reynolds and J. Postel, RFC 1700.

[7] Protocol Identification in the Network Layer, ISO/IEC TR 9577:1990.

Acknowledgements

We would like to thank Juha Heinenan of Telecom Finland and Ramesh Govidan of ISI for their work on NBMA ARP and the original NHRP draft, which served as the basis for this work. John Burnett of Adaptive, Dennis Ferguson of ANS, Joel Halpern of Newbridge, Paul Francis of NTT, and Tony Li and Bruce Cole of cisco should also be acknowledged for comments and suggestions that improved this work substantially. We would also like to thank the members of the Routing Over Large Clouds working group of the IETF, whose review and discussion of this document have been invaluable.

Authors' Addresses

Dave Katz
cisco Systems
170 W. Tasman Dr.
San Jose, CA 95134 USA

Phone: +1 408 526 8284
Email: dkatz@cisco.com

David Piscitello
Core Competence
1620 Tuckerstown Road
Dresher, PA 19025 USA

Phone: +1 215 830 0692
Email: dave@corecom.com