(from http://cell-relay.indiana.edu/mhonarc/rolc/1994-Jul/msg00003.html)

Routing over Large Clouds Working Group
Juha Heinanen
INTERNET-DRAFT
(Telecom Finland)
<draft-ietf-rolc-nhrp-XX.txt>
Ramesh Govindan

(Bellcore)
Dave Katz
(cisco Systems)

July 20, 1994

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 terminal (host or router) connected to a Non-Broadcast, Multi-Access link layer (NBMA) network to find out the IP and NBMA addresses of the "NBMA next hop" towards a destination terminal. The NBMA next hop is the destination terminal itself, if the destination is connected to the NBMA network. Otherwise, it is the egress router from the NBMA network that is "nearest" to the destination terminal. Although this document focuses on NHRP in the

context of IP, the technique is applicable to other network layer protocols as well.

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

1. Introduction

The NBMA Next Hop Resolution Protocol (NHRP) allows a source terminal (a host or router), wishing to communicate over a Non-Broadcast, Multi-Access link layer (NBMA) network, to find out the IP and NBMA addresses of the "NBMA next hop" towards a destination terminal. The NBMA next hop is the destination terminal itself, if the destination is connected to the NBMA network. Otherwise, it is the egress router (out of the NBMA network) nearest to the destination terminal.

Hop-by-hop IP routing may not be sufficient to resolve the "NBMA next hop" towards the destination terminal. An NBMA network may, in general, consist of multiple logically independent IP subnets (LISs, [4]); IP routing would only resolve the next hop LIS towards the destination terminal.

Once the NBMA next hop has been resolved, the source may either start sending IP packets to the destination (in a connectionless NBMA network such as SMDS) or may first establish a connection to the destination with the desired bandwidth and QOS characteristics (in a connection oriented NBMA network such as ATM).

An NBMA network can be non-broadcast either because it technically doesn't support broadcasting (e.g. an X.25 network) or because broadcasting is not feasible for one reason or another (e.g. an SMDS broadcast group or an extended Ethernet would be too large).

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 network, allowing for straightforward implementations in NBMA terminals. Optional services extend this functionality to include loop detection, sanity checks, diagnostics, and fallback capabilities, providing improved robustness and functionality.

NHRP supports both a server-based style of deployment and a ubiquitous "fabric". The server-based approach requires a smaller number of machines to support NHRP, but requires significantly more configuration.

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 network may be

partitioned into several disjoint logical NBMA networks (discussed later in this section); NHSs cooperatively resolve the NBMA next hop within their logical NBMA network. Unless otherwise specified, we use NBMA network to mean logical NBMA network.

S first determines the next hop to D through normal routing processes. If this next hop is reachable through its NBMA interface, S emits an NHRP request containing the source and destination IP addresses and QOS information.

Placed within the NBMA network 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 network.

When receiving an NHRP request, the NHS checks if it "serves" D. If so, the NHS resolves D's NBMA address, using information gleaned from NHRP Register packets, or mechanisms beyond the scope of this document (examples of such mechanisms include ARP [2, 3] and pre-configured tables). The NHS then generates a positive NHRP reply on its behalf. The reply contains the next hop IP and NBMA address for D and is sent back to S (note that if D is not on the NBMA network, the next hop IP address will be that of the egress router). NHRP replies usually traverse the same sequence of NHSs as the NHRP request (in reverse order, of course).

If the NHS does not serve D, it forwards the NHRP request to another NHS and the process is repeated. Mechanisms for determining how to forward the NHRP request are discussed below.

If the determination is made that D's next hop cannot be resolved, a negative reply is returned.

An NHS receiving an NHRP reply may cache the NBMA next hop information contained therein. To a subsequent NHRP request, this NHS might respond with the cached, non-authoritative, NBMA next hop or with cached negative information. If a communication attempt based on non-authoritative information fails, a source terminal can choose to send an authoritative NHRP request. NHSs never respond to authoritative NHRP requests with cached information.

NHRP requests and replies never cross the borders of a logical NBMA network. Thus, IP traffic out of and into a logical NBMA network always traverses an IP router at its border. Network layer filtering can then be implemented at these border routers.

NHRP optionally provides a mechanism to aggregate NBMA next hop information in NHS caches. Suppose that router X is the NBMA next

hop from S to D. Suppose further that X is an egress router for all terminals sharing an IP address prefix with D. When an NHRP reply is generated in response to a request, the responder may augment the IP address of D with a mask defining this prefix. The prefix to egress router mapping in the reply is cached in all NHSs on the path of the reply. A subsequent (non-authoritative) NHRP request for some destination that shares an IP address prefix with D can be satisfied with this cached information.

To dynamically detect link-layer filtering in NBMA networks, as well as to provide loop detection and diagnostic capabilities, NHRP optionally incorporates a "Route Record" in requests and replies. This Route record contains the network (and link 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 terminal is unable to open a connection to the responder, it may attempt to do so successively with other link layer addresses in the Route Record until it succeeds. This approach can find the optimal best hop in the presence of link-layer filtering.

3. Modes of Deployment

NHRP supports two deployment modes, which impact configuration and deployment complexity: "server" and "fabric" modes. These 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 network 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.

Server Mode

In "server" mode, the expectation is that a small number of NHSs will be fielded in an NBMA network. This may be appropriate in networks containing routers that do not support NHRP, or networks 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 path taken by NHRP requests has little to do with the path taken by IP data packets routed to the destination.

Server mode uses static configuration of NHS identity. The source terminal 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 NHSs NBMA address must also 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 and the identities of the destinations that each of them serves. (This static configuration requirement tends to limit such deployments to very small number of NHSs.)

The NHRP packet's destination IP address is set by the source terminal to the 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.

Fabric Mode

In "fabric" mode, it is expected that NHRP-capable routers are ubiquitous throughout the NBMA network. In particular, it is expected that an NHS serving a particular destination is guaranteed to 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 network are destinations beyond them, and that hosts on the NBMA network are served by routers that double as NHSs.

Fabric mode leverages a routed infrastructure that "overlays" the NBMA network. The source terminal 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 it hits a router that is acting as an NHS serving the destination, which generates the NHRP reply.

If the source terminal 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 network is running). If the source terminal is a router, the destination IP address may be set either to the next-hop router or to the ultimate destination. Each NHS/router examines the packet on its way toward the destination, optionally modifying it on the way (such as updating the Forward Record option). 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 hit the router/NHS that serves the destination (which will return a positive NHRP reply) or it will hit a router

that has no route to the destination (which will return a negative NHRP reply).

The procedural difference between server mode and fabric mode boils down to deciding how to update the destination address in the IP packet carrying the NHRP request (see Protocol Operation).

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

4. Configuration

Terminals

To participate in NHRP, a terminal connected to an NBMA network should to be configured with the IP address(es) of its NHS(s). These NHS(s) may be physically located on the terminal's default or peer routers, so their addresses may be obtained from the terminal's IP forwarding table. If the terminal is attached to several link layer networks (including logical NBMA networks), it should also be configured to receive routing information from its NHS(s) and peer routers so that the terminal can determine which IP networks are reachable through which link layer networks.

Next Hop Servers

An NHS is configured with a set of IP address prefixes that correspond to the IP addresses of the terminals it is serving. If a served terminal is attached to several link layer networks, the NHS may also need to be configured to advertise routing information to such terminals.

If an NHS is acting as an egress router for terminals connected to other link layer networks than the NBMA network, the NHS must, in addition to the above, be configured to exchange routing information between the NBMA network and these other link layer networks.

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

The NBMA addresses of the terminals served by the NHS will typically be learned via NHRP Register packets; manual configuration is also possible, but unwieldy.

5. Packet Formats

NHRP packets are carried in IP packets as protocol type <TBD>. This section describes the format of NHRP packets.

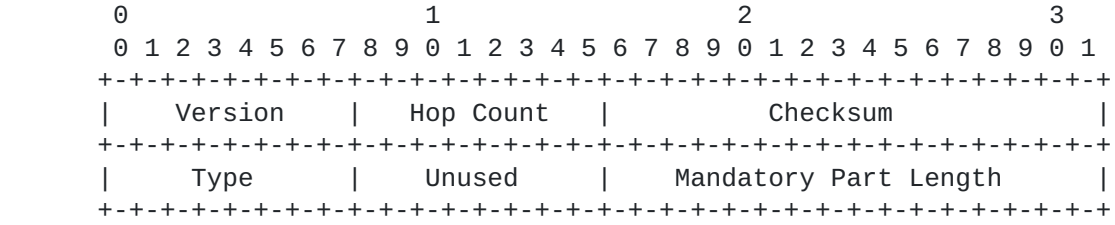
An NHRP packet consists of a Fixed Header, a Mandatory Part, and an Options Part. The fixed header is common to all NHRP packet types. The Mandatory Part varies depending on packet type, but must be present. The Options Part also varies depending on packet type, and need not be present.

The length of the Fixed Header is, of course, fixed. The length of the Mandatory Part is carried in the Fixed Header. The length of the Options Part is implied by the total packet length.

Note that since the lengths of all fields are self-encoding, it is permissible to pad the Mandatory and Options parts with trailing arbitrary numbers of trailing zero octets to achieve any desired alignment. Note however that any padding in the Mandatory Part must be included in the Mandatory Part Length.

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).

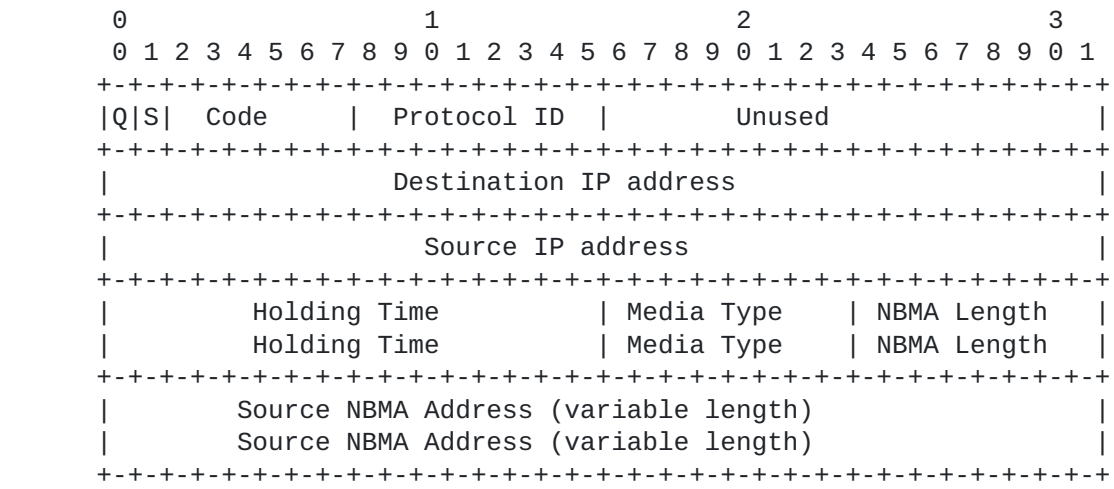
Type

The NHRP packet type (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)

Code
A response to an NHRP request may contain cached information. If an authoritative answer is desired, then code 2 (NHRP request for authoritative information) should be used. Otherwise, a code value of 1 (NHRP request) should be used.

Protocol ID
Specifies the network layer protocol for which we are obtaining routing information. This value also qualifies the structure of the remainder of the Mandatory Part. For IP, the Protocol ID is hexadecimal CC (decimal 204). Protocol ID values for other network layer protocols are for future study.

Destination and Source IP Addresses

Respectively, these are the IP addresses of the terminal for which

the NBMA next hop is desired, and the NHRP request initiator.

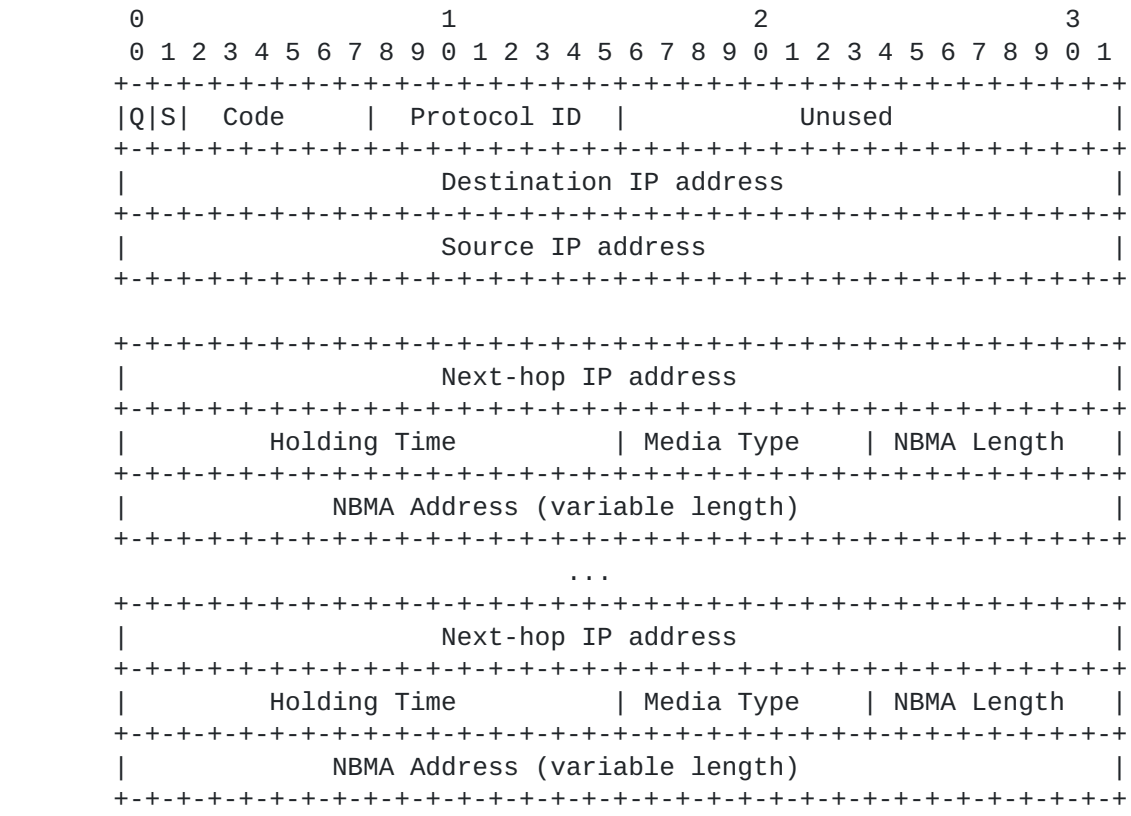
Source Holding Time, Media Type, NBMA Length, and NBMA Address
The Holding Time field specifies the number of seconds for which the source NBMA information is considered to be valid. Cached information shall be discarded when the holding time expires.

The Media Type field specifies the type of NBMA media (qualifying the NBMA address). Possible media types are <TBD>. [Note: I believe there is ongoing work on canonical identifiers for media types. If true, this document should be aligned with that work. --Dave]

The NBMA length field is the length of the NBMA address of the source terminal in bits. The NBMA address 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.

Code
NHRP replies may be positive or negative. An NHRP positive, non-authoritative reply carries a code of 1, while a positive, authoritative reply carries a code of 2. An NHRP negative, non-authoritative reply carries a code of 3 and a negative, authoritative reply carries a code of 4. 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.

Destination IP Address
The address of the target terminal of the corresponding NHRP Request.

Source IP Address

The address of the initiator of the corresponding NHRP Request.

Next-hop IP Address, Holding Time, Media Type, NBMA Length, and NBMA Address

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 network, or the egress router if not.

The Holding Time field specifies the number of seconds for which the information carried in the reply is considered to be valid. Cached information shall be discarded when the holding time expires.

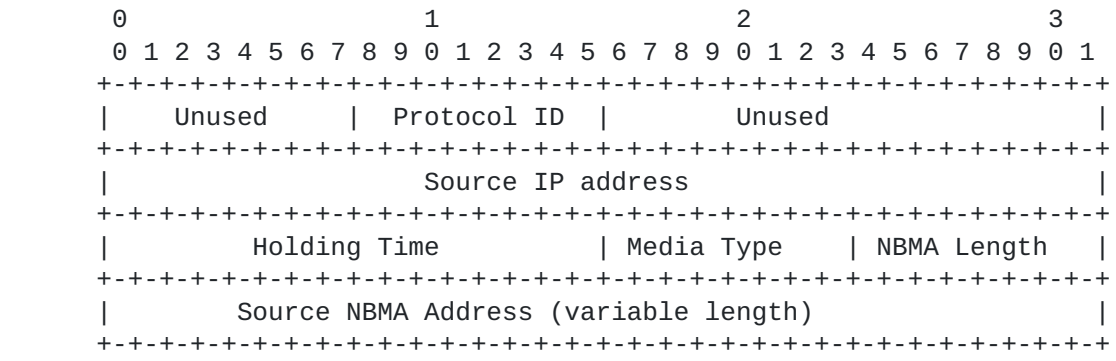
The Media Type field specifies the type of NBMA media (qualifying the NBMA address). Possible media types are <TBD>. [Note: I believe there is ongoing work on canonical identifiers for media types. If true, this document should be aligned with that work. --Dave]

The NBMA length field is the length of the NBMA address of the destination terminal in bits. The NBMA address itself is zero-filled to the nearest 32-bit boundary. Negative replies do not carry the media type, NBMA length, or NBMA address fields.

There may be multiple IP/NBMA addresses returned in the reply (as implied by the Mandatory Part Length). These addresses should be considered to be equivalent. The same next-hop IP address may be associated with multiple NBMA addresses. The choice of which address to use is a local matter, although it is encouraged that load-splitting be performed over the addresses, and that the alternative addresses be used in case of connectivity failure in the NBMA network (such as a failed call attempt in connection-oriented NBMA networks).

5.4 NHRP Register

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



Protocol ID
Specifies the network layer protocol for which we are obtaining routing information. This value also qualifies the structure of the remainder of the Mandatory Part. For IP, the Protocol ID is hexadecimal CC (decimal 204). Protocol ID values for other network layer protocols are for future study.

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

Source Holding Time, Media 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 Media Type field specifies the type of NBMA media (qualifying the NBMA address). Possible media types are <TBD>. [Note: I believe there is ongoing work on canonical identifiers for media types. If true, this document should be aligned with that work. --Dave]

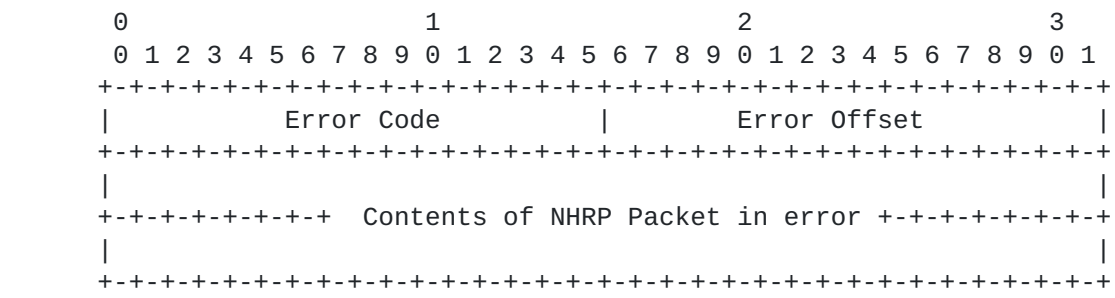
The NBMA length field is the length of the NBMA address of the source terminal in bits. The NBMA address itself is zero-filled to the nearest 32-bit boundary.

This packet is used to register a terminal'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 terminals that they serve.

It is possible that a misconfigured terminal 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 will reply with an Error Indication of type Can't Serve This Address.

5.5 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 4. 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 Option
2 - Network ID Mismatch
3 - NHRP Loop Detected
4 - Can't Serve This Address
4 - Can't Serve This Address

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). In other words, the Error Indication shall be sent directly to the initiator, rather than traversing each NHS along the way.

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 it be possible for more than one Error Indication packet to be generated for a single NHRP packet.

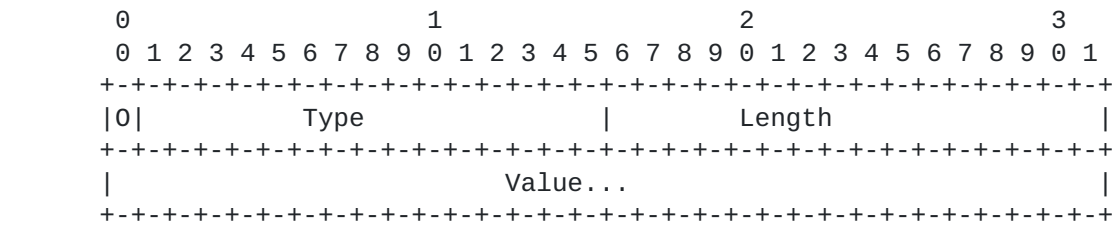
5.6 Options Part

The Options Part, if present, carries one or more options in {Type, Length, Value} triplets. Options are only present in a Reply if they were present in the corresponding Request; therefore, minimal NHRP

terminal implementations that do not act as an NHS and do not

transmit options need not be able to receive them. Such an implementation that receives a packet with options shall return an Error Indication of type Unrecognized Option.

Options have the following format:



O
"Optional." If set, and the NHS does not recognize the type code, the option 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 option type code (see below). The option type is not qualified by the Optional bit, but is orthogonal to it.

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

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

Options may occur in any order, but any particular option type may occur only once in an NHRP packet.

The Optional bit provides for a means to extend the option set. If it is clear, the NHRP request cannot be satisfied if the option is unrecognized, so the responder must return an Error Indication of type Unrecognized Option. If set, the option can be safely ignored. In this case, the offending option should simply be returned unchanged in the NHRP Reply.

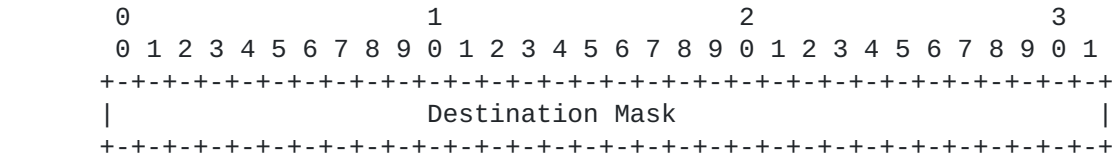
If a transit NHS (one which is not going to generate a reply) detects an unrecognized option, it shall ignore the option, and if the Optional bit is clear, must not cache the information (in the case of a reply) and must not identify itself as an egress router (in the Forward Record or Reverse Record options). Effectively, this means that a transit NHS that doesn't understand an option with the

Optional bit clear must not participate in any way in the protocol exchange, other than acting as a forwarding agent for the request.

5.6.1 Destination Mask Option

Optional = 0
Type = 1
Length = 4

This option 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 destination IP address in the bit positions for which the mask has a one bit are part of the equivalence class.

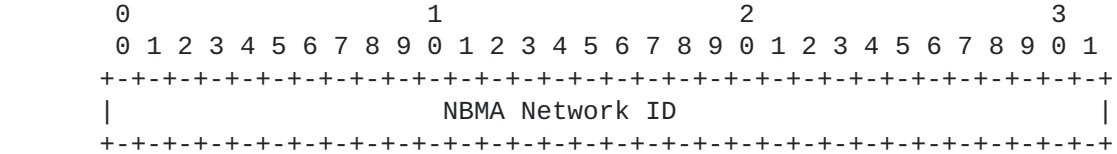


If an initiator would like to receive this equivalence information, it shall add this option to the NHRP Request with a value of 255.255.255.255. The responder shall copy the option to the NHRP Reply and modify the mask appropriately.

5.6.2 NBMA Network ID Option

Optional = 0
Type = 2
Length = 4

This option is used to carry an identifier for the NBMA network. This can be used as a validity check to insure that the request does not leave a particular NBMA network. The option 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 for the NBMA network. NHS fills in the field with the identifier for the NBMA network.



The identifier consists of a 32 bit globally unique value assigned to the NBMA network. This value should be chosen from the IP address space administered by the operators of the NBMA network. This value

is used for identification only, not for routing or any other purpose.

Each NHS processing an NHRP Request shall verify this value. If the value does not match the NHS's NBMA Network ID, the NHS shall return an Error Indication of type "Network ID Mismatch" and discard the NHRP Request.

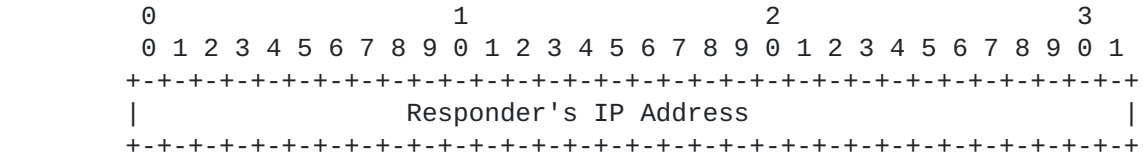
When an NHS is building an NHRP Reply and the NBMA Network ID option is present in the NHRP Request, the NBMA Network ID option shall be copied from the Request to the Reply.

Each NHS processing an NHRP Reply shall verify the value carried in the NBMA Network ID option, if present. If the value does not match the NHS's NBMA Network ID, the NHS shall return an Error Indication of type "Network ID Mismatch" and discard the NHRP Reply.

5.6.3 Responder Address Option

Optional = 0
Type = 3
Length = 4

This option is used to determine the IP address of the NHRP Responder, that is, the entity that generates the NHRP Reply packet.



If a requestor desires this information, it shall include this option, 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 option, it shall include this option, 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 options.

If an NHRP Reply packet being forwarded by an NHS contains the IP address of that NHS in the Responder Address Option, 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 option (differentiating it from the address in the Next-hop field).

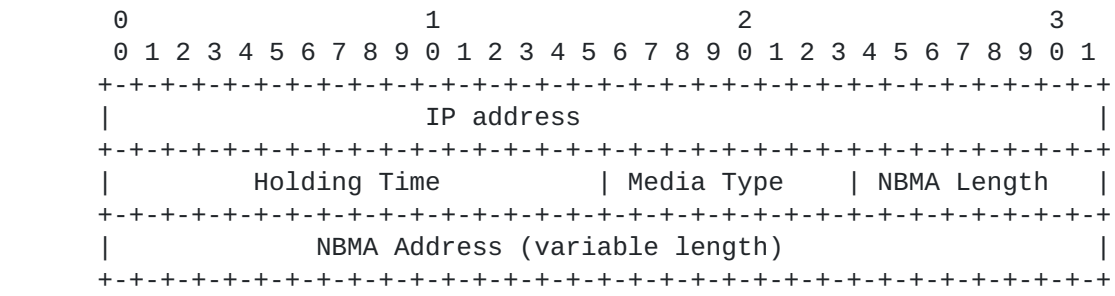
5.6.4 NHRP Forward NHS Record Option

Optional = 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 an NHS element containing its IP address to this option.

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 NHS 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 terminal chooses to use this information as a next-hop entry, it may not be used once the holding timer expires.

Media Type, NBMA Length, and NBMA Address
The Media Type field specifies the type of NBMA media (qualifying the NBMA address). Possible media types are TBD.

The NBMA length field is the length of the NBMA address of the destination terminal 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 option with a length of zero.

Each NHS shall append an appropriate NHS element to this option when processing an NHRP Request. The option length field and NHRP checksum shall be adjusted as necessary.

The last element shall contain information about the responder.

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 options.

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

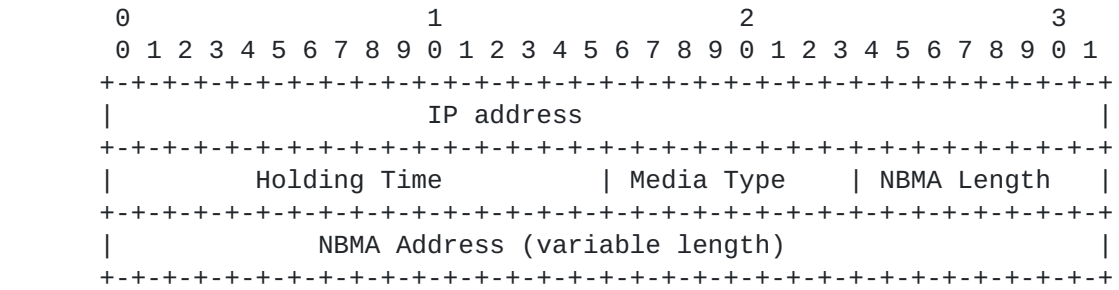
5.6.5 NHRP Reverse NHS Record Option

Optional = 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 an NHS element containing its IP address to this option.

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 NHS 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 terminal chooses to use this information as a next-hop entry, it may not be used once the holding timer expires.

Media Type, NBMA Length, and NBMA Address
The Media Type field specifies the type of NBMA media (qualifying the NBMA address). Possible media types are TBD.

The NBMA length field is the length of the NBMA address of the destination terminal 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 option with a length of zero.

Each NHS shall append an appropriate NHS element to this option when processing an NHRP Reply. The option length field and NHRP checksum shall be adjusted as necessary.

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 options.

If an NHRP Reply packet being forwarded by an NHS contains the IP address of that NHS in the Reverse NHS Record Option, 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 option may be used to disambiguate the set of NHSs that actually processed the reply.

5.6.6 NHRP QoS Option

Optional = 0
Type = 6
Length = variable

The NHRP QoS Option 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 option are TBD.

6. Discussion

The result of an NHRP request depends on how routing is configured among the NHSs of an NBMA network. If the destination terminal is directly connected to the NBMA network and the NHSs always prefer directly connected to the NBMA network and the NHSs always prefer NBMA routes over routes via other link layer networks, the NHRP replies always return the NBMA address of the destination terminal itself rather than the NBMA address of some egress router. For destinations outside the NBMA network, egress routers and routers in the other link layer networks should exchange routing information so that the optimal egress router is always found.

When the NBMA next hop towards a destination is not the destination terminal 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 terminal will need to periodically reissue the NHRP Request at a minimum just prior to the expiration of the holding timer, and most likely more aggressively than that. Alternatively, the source can be configured to receive routing information from its NHSs. 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 terminals 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.)

In addition to NHSs, an NBMA terminal could also be associated with one or more regular routers that could act as "connectionless servers" for the terminal. Then the terminal could choose to resolve the NBMA next hop or just send the IP packets to one of the terminal's 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.

NHRP supports portability of NBMA terminals. A terminal can be moved
anywhere within the NBMA network and still keep its original IP

address as long as its NHS(s) remain the same. Requests for authoritative information will always return the correct link layer address.

7. Protocol Operation

7.1 Router-to-Router Operation

In practice, the initiating and responding terminals 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. This situation can be avoided if there are no "back door" paths between the entry and egress router outside of the NBMA network, and can be ameliorated by periodically reissuing the NHRP request. If these conditions cannot be satisfied, the use of NHRP between routers is not recommended.

7.2 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 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 quietly 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.

Terminals

Terminals 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.

7.3 Pseudocode

TBD.

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, Internet Draft.

[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.

Acknowledgements

We would like to thank John Burnett of Adaptive, Dennis Ferguson of ANS, Joel Halpern of Network Systems, and Paul Francis of Bellcore for their valuable insight and comments to earlier versions of this draft.

Authors' Addresses

Juha Heinanen
Telecom Finland,
PO Box 228,
SF-33101 Tampere,
Finland

Phone: +358 49 500 958
Email: Juha.Heinanen@datanet.tele.fi

Ramesh Govindan
Bell Communications Research
MRE 2P-341, 445 South Street
Morristown, NJ 07960

Phone: +1 201 829 4406
Email: rxg@thumper.bellcore.com

Dave Katz
cisco Systems
1525 O'Brien Dr.
Menlo Park, CA 94025 USA

Phone: +1 415 688 8284
Email: dkatz@cisco.com