Network Working Group
Request for Comments: 1735
Category: Experimental

J. Heinanen
Telecom Finland
R. Govindan
ISI

December 1994

NBMA Address Resolution Protocol (NARP)

Status of this Memo

This memo defines an Experimental Protocol for the Internet community. This memo does not specify an Internet standard of any kind. Discussion and suggestions for improvement are requested. Distribution of this memo is unlimited.

IESG Note:

Note that the work contained in this memo does not describe an Internet standard. This work represents an early stage in the ongoing efforts to resolve direct communication over NBMA subnets. It is a suitable experimental protocol for early deployment. It is expect that it will be superceded by other work being developed within the IETF.

Abstract

This document describes the NBMA Address Resolution Protocol (NARP). NARP 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 NBMA addresses of the a destination terminal provided that the destination terminal is connected to the same NBMA network. Although this document focuses on NARP in the context of IP, the technique is applicable to other network layer protocols as well. This RFC is a product of the Routing over Large Clouds Working Group of the IETF.

1. Introduction

The NBMA Address Resolution Protocol (NARP) 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 NBMA addresses of a destination terminal if the destination terminal is connected to the same NBMA network as the source.

A conventional address resolution protocol, such as ARP [1, 2] for IP, may not be sufficient to resolve the NBMA address of the destination terminal, since it only applies to terminals belonging to the same IP subnetwork, whereas an NBMA network can consist of multiple logically independent IP subnets (LISs, [3]).

Once the NBMA address of the destination terminal is 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).

2. Protocol Overview

In this section, we briefly describe how a source S uses NARP to determine the NBMA address of a destination D or to find out that such an address doesn't exist. S first checks if the destination terminal belongs to the same IP subnetwork as S itself. If so, S resolves the NBMA address of D using conventional means, such as ARP [1, 2] or preconfigured tables. If D resides in another subnetwork, S formulates a NARP request containing the source and destination IP addresses. S then forwards the request to an entity called the "NBMA ARP Server" (NAS).

For administrative and policy reasons, a physical NBMA network may be partitioned into several disjoint logical NBMA networks. NASs cooperatively resolve the NBMA next hop within their logical NBMA network. In the following we'll always use the term "NBMA network" to mean a logical NBMA network. If S is connected to several NBMA networks, it should have at least one NAS in each of them. In order to know which NAS(s) to query for which destination addresses, a multi-homed S should also be configured to receive reachability information from its NASs.

Each NAS "serves" a pre-configured set of terminals and peers with a pre-configured set of NASs, which all belong to the same NBMA network. A NAS may also peer with routers outside the served NBMA. A NAS exchanges reachability information with its peers (and possibly with the terminals it serves) using regular routing protocols. This exchange is used to construct a forwarding table in every NAS. The forwarding table determines the next hop NAS towards the NARP request's destination or a next hop router outside the NBMA.

After receiving a NARP request, the NAS checks if it "serves" D. If so, the NAS resolves D's NBMA address, using mechanisms beyond the scope of this document (examples of such mechanisms include ARP [1, 2] and pre-configured tables). The NAS then either forwards the NARP request to D or generates a positive NARP reply on its behalf. The reply contains D's IP and NBMA address and is sent back to S. NARP replies usually traverse the same sequence of NASs as the NARP request (in reverse order, of course).

If the NAS does not serve D, it extracts from its forwarding table the next hop towards D. If the next hop is a peer NAS, it forwards the NARP request to the next hop. If the next hop is a peer router outside the served NBMA or if no such next hop entry is found, the NAS generates a negative NARP reply.

A NAS receiving a NARP reply may cache the NBMA address information contained therein. If a subsequent NARP request for the same target address does not desire an authorative reply, a caching NAS can then respond with the cached non-authoritative NBMA address or with cached negative information. A well behaving terminal should always first accept a non-authoritative reply. Only if communication attempt based on the non-authoritative information fails, the terminal can choose to issue another request this time asking for an authoritative reply.

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

3. Configuration

Terminals

To participate in NARP, a terminal connected to an NBMA network should to be configured with the IP address(es) of its NAS(s). If the terminal is attached to several NBMA networks, it should also be configured to receive reachability information from its NAS(s) so that it can determine, which IP destinations are reachable through which NBMA networks.

NBMA ARP Servers

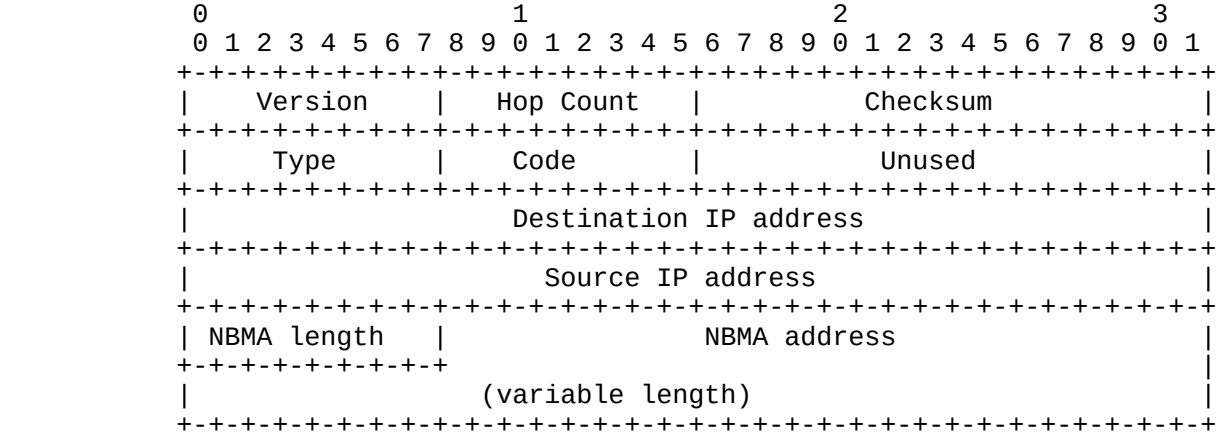
A NAS is configured with a set of IP address prefixes that correspond to the IP addresses of the terminals it is serving. Moreover, the NAS must be configured to exchange reachability information with its peer NASs (if any). In addition, the NAS may be configured to exchange reachability information with routers

outside the served NBMA. And finally, if a served terminal is attached to several NBMA networks, the NAS may need to be configured to send reachability information to such a terminal.

4. Packet Formats

NARP requests and replies are carried in IP packets as protocol type 54. This section describes the packet formats of NARP requests and replies:

NARP Request



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

Hop Count
The Hop count indicates the maximum number of NASs that a request or reply is allowed to traverse before being discarded.

Checksum
The standard IP checksum over the entire NARP packet (starting with the fixed header).

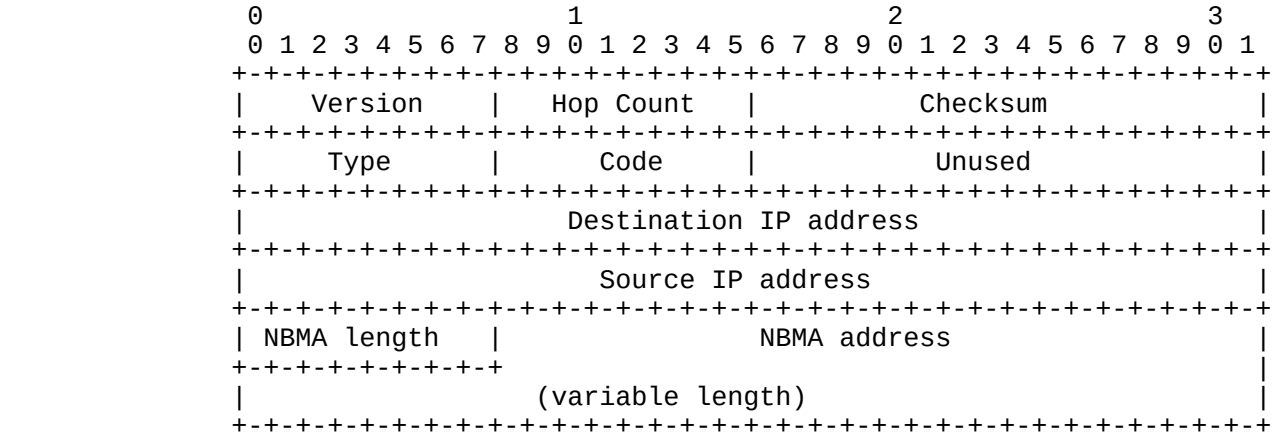
Type
The NARP packet type. The NARP Request has a Type code 1.

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

Source and Destination IP Addresses
Respectively, these are the IP addresses of the NARP requestor and the target terminal for which the NBMA address is desired.

NBMA Length and NBMA Address
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.

NARP Reply



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

Hop Count
The Hop count indicates the maximum number of NASs that a request or reply is allowed to traverse before being discarded.

Checksum
The standard IP checksum over the entire NARP packet (starting with the fixed header).

Type
The NARP packet type. The NARP Reply has a Type code 2.

Code
NARP replies may be positive or negative. A Positive, Non-authoritative Reply carries a code of 1, while a Positive, Authoritative Reply carries a code of 2. A Negative, Non-authoritative Reply carries a code of 3 and a Negative, Authoritative reply carries a code of 4.

The general rule is that a NAS should not reply to an NARP request for authoritative information with cached information, but may do so for an NARP request. A NAS implementation is allowed to relax this rule and return non-authoritative information even in case authorative was desired if the NAS becomes heavily loaded and the cached information is very recently updated.

Source and Destination IP Address
Respectively, these are the IP addresses of the NARP requestor and the target terminal for which the NBMA address is desired.

NBMA Length and NBMA Address
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 NBMA length or the NBMA address field.

A NAS may cache NBMA replies.

5. Protocol Operation

The external behavior of a NAS may be described in terms of two procedures (processRequest and processReply) operating on two tables (forwardingTable and cacheTable). In an actual implementation, the code and data structures may be realized differently.

Each NAS has a forwardingTable consisting of entries with the fields:

<networkLayerAddrPrefix, type, outIf, outIfAddr>

The networkLayerAddrPrefix field identifies a set of IP addresses known to the NAS. It consists of two subfields <ipAddr, mask>.

The type field indicates the type of the networkLayerAddrPrefix. The possible values are:

- locallyServed: The NAS is itself serving the networkLayerAddrPrefix. The outIf field denotes the NBMA interface via which the served terminals can be reached and the outIfAddr field has no meaning. Such a forwardingTable entry has been created by manual configuration.

- nasLearned: The NAS has learned about the networkLayerAddrPrefix from another NAS. The outIf and outIfAddr fields, respectively, denote the NBMA interface and IP address of this next hop NAS. Such a forwardingTable entry is a result of network layer address prefix information exchange with one of the NAS' peer NASs.

- externallyLearned: The NAS has learned about the networkLayerAddrPrefix from a peer router outside the served NBMA. The outIf and outIfAddr fields, respectively, denote the NBMA interface and IP address of this next hop NAS. Such a forwardingTable entry is a result of network layer address prefix information exchange with one of the NAS' peer routers.

The protocol used to exchange networkLayerAddrPrefix information among the NASs can be any regular IP intra-domain or inter-domain routing protocol.

In addition to the forwardingTable, each NAS has an NARP cacheTable consisting of entries with the fields:

<networkLayerAddr, nbmaAddr, timeStamp>

The entries in the cacheTable are learned from NARP replies traversing the NAS. In case of a negative cache entry the nbmaAddr is empty. The timeStamp field records the time when the cacheTable entry has been created or updated. It is used to determine if an entry is a very recent one and to age old entries after a certain hold period.

The following pseudocode defines how NBMA NARP requests and replies are processed by an NAS.

```
procedure processRequest(request);
  let bestMatch == matchForwardingTable(request.dIPa) do
     if bestMatch then
        if bestMatch.type == locallyServed then
           let nbmaAddr == arp(request.dIPa) do
              if nbmaAddr then
                 genPosAuthReply(request.sIPa, request.dIPa, nbmaAddr)
              else
                 genNegAuthReply(request.sIPa, request.dIPa)
              end
           end
        elseif bestMatch.type == nasLearned then
           if not requestForAuthInfo?(request) or
                 realBusyRightNow?() then
              let cacheMatch == matchCacheTable(request.dIPa) do
                 if cacheMatch and
                       (not requestForAuthInfo?(request) or
                          realRecentCacheEntry?(cacheMatch)) then
                    if cacheMatch.nbmaAddr == EMPTY then
                       genNegNonAuthReply(request.sIPa, request.dIPa)
                    else
                       genPosNonAuthReply(request.sIPa, request.dIPa,
```

```
                        cacheMatch.nbmaAddr)
                  end
               else /* no cache match */
                  forwardRequest(request, bestMatch.OutIf,
                     bestMatch.OutIfAddr)
               end
            end
         else /* request for authoritative information */
            forwardRequest(request, bestMatch.OutIf,
               bestMatch.OutIfAddr)
         end
      else /* bestMatch.type == externallyLearned */
         genNegAuthReply(request.sIPa, request.dIPa)
      end
   else /* no match in forwardingTable */
      genNegAuthReply(request.sIPa, request.dIPa)
   end
 end
end

procedure processReply(reply);
 addCacheTableEntry(reply.dIPa, reply.nbmaAddr, currentTime);
 if reply.sIPa == selfIpAddr then
    /* reply is to the NAS itself */
 else
    let bestMatch == matchForwardingTable(reply.sIPa) do
       if bestMatch then
          forwardReply(reply, bestMatch.outIf, bestMatch.outIfAddr)
       end
    end
 end
end
```

The semantics of the procedures used in the pseudocode are explained below.

matchForwardingTable(ipAddress) returns the forwardingTable entry whose networkLayerAddrPrefix field is the longest match for ipAddress or FALSE if no match is found.

arp(ipAddress) resolves the NBMA address corresponding to ipAddress. It returns FALSE if the resolution fails.

genPosAuthReply(sourceIpAddr, destIpAddr, destNbmaAddr) and genPosNonAuthReply(sourceIpAddr, destIpAddr, destNbmaAddr) generate a positive, authoritative and non-authoritative reply with sourceIpAddr, destIpAddr, and destNbmaAddr in Source IP address, Destination IP address, and NBMA Address fields, respectively.

genNegAuthReply(sourceIpAddr, destIpAddr) and genNegNonAuthReply(sourceIpAddr, destIpAddr) respectively generate a negative, authoritative and non-authoritative reply with sourceIpAddr and destIpAddr in Source IP address and Destination IP address fields, respectively.

requestForAuthInfo?(request) tests if request is a Request for authoritative information.

realBusyRightNow?() returns TRUE if the NAS is heavily loaded and FALSE otherwise.

realRecentCacheEntry?(cacheTableEntry) returns TRUE if the cacheTableEntry is very recently updated and FALSE otherwise.

matchCacheTable(ipAddr) returns a cacheTable entry whose networkLayerAddr field is equal to ipAddr or FALSE if no match is found.

forwardRequest(request, interface, ipAddr) decrements the Hop count field of request, recomputes the NARP Checksum field, and forwards request to ipAddr of interface provided that the value of the Hop count field remains positive.

addCacheTableEntry(ipAddr, nbmaAddr, time) adds a new entry to the cacheTable or overwrites an existing entry whose networkLayerAddr field is equal to ipAddr.

forwardReply(reply, interface, ipAddr) decrements the Hop count field of request, recomputes the NARP Checksum field, and forwards reply to ipAddr of interface provided that the value of the Hop count field remains positive.

Like NASs, each NBMA terminal has a forwardingTable and a cacheTable. The forwardingTable is either manually configured or filled via reachability information exchange with the terminal’s NASs or peer routers.

When the terminal wishes to find out the NBMA address of a particular destination terminal, it first checks if a matching entry is found in the forwardingTable. If not, the destination is unreachable and the terminal gives up. If a forwardingTable entry is found, and if the next hop belongs to one of the terminal’s NASs, the terminal next consults its cacheTable to obtain the NBMA address. If no cache match is found, the terminal generates a NARP request to the next hop NAS. If the reply to the NARP request is positive, the terminal learns the NBMA address and updates its cacheTable with the new information.

6. Discussion

The NARP semantics resembles closely the ATMARP semantics described in [2]. The only actual differences are:

- NARP requests and replies include a hop count to prevent them from looping forever in case of misconfigured NAS routing.

- NARP request and replies distinguish between authoritative and non-authoritative information.

In order to keep the NBMA terminals as simple as possible, it would be desirable to extend the the ATMARP protocol a little further so that it could be also used as the terminal-NAS protocol. This could be easily accomplished just by adding three new operation codes to ATMARP to cover the different kinds of queries and responses. NARP would then become the NAS-NAS protocol. Finally, if the NASs are co-located with the "classical" ATM ARP servers, the terminals would not need to make any distinction between between local and foreign IP subnetworks.

The NASs can also act as "connectionless servers" for the terminal by advertizing to it all destinations no matter if they are inside or outside the served NBMA. Then, the terminal could choose either to try to resolve the NBMA address of the destination or just to send the IP packets to the NAS. The latter option may be desirable if communication with the destination is short-lived and/or doesn't require much network resources.

NARP 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 NAS(s) remain the same. Requests for authoritative information will always return the correct NBMA address.

References

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

[2] Laubach, M., "Classical IP and ARP over ATM", RFC 1577, Hewlett-Packard Laboratories, January 1994.

[3] Piscitello, D., and J. Lawrence, "Transmission of IP Datagrams over the SMDS Service, RFC 1209, Bell Communications Research, March 1991.

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.

Security Considerations

Security issues are not discussed in this memo.

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
USC/Information Sciences Institute
4676 Admiralty Way
Marina del Rey, CA 90292

Phone: +1 310-822-1511
EMail: govindan@isi.edu