Network Working Group
Request for Comments: 2583
Category: Informational

R. Carlson
ANL
L. Winkler
ANL

May 1999

Guidelines for Next Hop Client (NHC) Developers

Status of this Memo

This memo provides information for the Internet community. It does not specify an Internet standard of any kind. Distribution of this memo is unlimited.

Copyright Notice

1. Abstract

This document provides guidelines for developers of the Next Hop Resolution Protocol Clients (NHC). It assumes that the clients are directly connected to an ATM based NBMA network. The same principles will apply to clients connected to other types of NBMA networks. The intent is to define the interaction between the NHC code and the TCP/IP protocol stack of the local host operating system. The NHC is capable of sending NHRP requests to a Next Hop Resolution Protocol Server (NHS) to resolve both inter and intra LIS addresses. The NHS reply may be positive (ACK) indicating a short-cut path is available or negative (NAK) indicating that a shortcut is not available and the routed path must be used. The NHC must cache (maintain state) for both the ACK and NAK replies in order to use the correct shortcut or routed path. The NAK reply must be cached to avoid making repeated requests to the NHS when the routed path is being used.

2. Overview

In the Classical IP over ATM model [1], an ATM attached host communicates with an ATMARP server to resolve IP to ATM address semantics. This model supports the concept of a Logical IP Subnet (LIS) with intra LIS communications using direct PVCs/SVCs and inter LIS communications using IP routers to forward packets. This model easily maps to the conventional LAN model of subnets and routers. The Next Hop Resolution Protocol (NHRP) [2] defines how the LIS model can be modified to allow direct ATM SVCs (shortcut paths) for inter LIS traffic. With NHRP, nodes directly attached to an ATM network can bypass the IP routers and establish a direct switched virtual

circuit to improve performance when needed.

The NHS code replaces the ATMARP code in the ATMARP server. Each NHS serves a set of destination client hosts and cooperates with other NHSs to resolve NHRP next hop requests within their own logical ATM network. The NHC to NHS and NHS to NHS protocol interactions are described in [2]. Other documents in the NHRP series define the general applicability [3] and the transition from ATMARP servers to NHSs [4].

The NHC code replaces the ATMARP code in the local workstations. This code will take the destination IP address and map it into the ATM End Station Address (AESA) for both intra and inter LIS destinations. The returned AESA will be stored in a local cache table. In addition to storing the positive replies, the NHC will need to store the negative replies to avoid making repeated NHS calls when using the routed path.

This document describes a base line method for caching the returned information. Other methods may be used as long as the same functionality is provided.

3. IP Processing

In the Classical IP LIS model [1] the TCP/IP protocol stack treats the ATM network as a simple data link layer protocol. When an application sends data using the Classical IP protocol, IP performs a routing table lookup to determine if the destination is reachable via a local interface or whether an intermediate router is the next hop to the IP destination.

If the destination is found to be local (e.g. in the same LIS as the source) the packet will be passed to the local ATM interface with the next hop IP address set to the destination nodes IP address. At this point the ATMARP table will be searched to determine the ATM Address of the destination node. If no ATMARP table entry is found an ATMARP request will be sent to the ATMARP server. This server can reply with a positive (ACK) or negative (NAK) answer depending on the current information it has in its cache. If an ACK is received the host’s local ATMARP table is filled in appropriately and the source is now able to send IP datagrams to the destination. If a NAK is returned, the calling application is notified of this error condition (e.g., ICMP destination unreachable).

If the destination is found to be remote (e.g., in a different LIS from the source) the IP address of the next hop router is extracted from the IP routing table and the ATM Address of this router is looked up in the ATMARP table. Since the router is in the same LIS

as the source node, the ATMARP procedure described above will find the correct ATM Address or the packet will be marked as undeliverable and the user application will be notified of the error.

The ATMARP service functions exactly as the existing ARP service provided on Ethernet broadcast networks. Since the ARP service will only try and resolve addresses for nodes that are in a single IP subnet, the ARP table only needs to keep positive answers. No state information is retained about failed mappings.

4. NHC Processing

In this section we briefly describe what is required in order for a host to take advantage of shortcuts through the ATM network. On the host, a NHC process initiates various NHRP requests in order to obtain access to the NHRP service. Within the ATM subnetwork, the ATMARP server is replaced with a NHS. As defined in [4] the NHS is required to respond to both ATMARP and NHRP Resolution requests. In the nodes wishing to take advantage of shortcut paths across the ATM subnetwork, the ATMARP client code must be replaced with NHC code. This allows the source node to ask for the ATM AESA of both local and remote nodes. Finally the source node must be modified to know when it should ask for the ATM AESA of a remote node and when the local LIS router should be used. These modifications are described in the remainder of this document.

The protocol processing described in [2] states a source may query a NHS for the ATM AESA of a destination node. However as is pointed out in [5], to achieve shortcut paths through the ATM network, it is not enough to simply replace the ATMARP client code with the NHC code. This is because the source host will never ask the NHS for the ATM AESA of a node in a remote LIS. When the source consults the IP routing table, it performs the local/remote test, before the NHC code is processed. As a result, the IP address of the next hop router will be used by the NHC instead of the IP address of the remote (inter LIS) host. The NHC code must ignore the result of the IP routing table lookup and perform its own local/remote test.

The NHC must perform the following functions:

1. Test to see if the destination node is ‘local’ to this LIS. If so use the existing ATMARP rules described in [1].
2. If not; send an NHRP message to the local NHS and attempt to setup a ‘shortcut’ path. If successful; save the IP to ATM AESA mapping in the local NHC cache.
3. If not successful; use the routed path and save this state in the NHC cache so future requests don’t test for a shortcut again.

4. Allow user application to override system default operation and explicitly request a shortcut or routed path for a flow.

It is required that this routed path state will be maintained in the same manner as the existing ATMARP service. That is a timer will be used to expire old information and some administrative function exists to manually delete data if needed.

5. Need for State

It is obvious that the IP to ATM AESA mappings should be maintained in a local cache to improve network performance. This soft state is maintained in today's ARP and ATMARP systems using timers to purge old or unused data. The NHC will maintain both inter and intra LIS IP to ATM Address mappings in the same manner. It may be less obvious that an NHC will also need to maintain this same soft state for inter LIS mappings using the routed path. If this state is not maintained, the source node will send requests to the NHS asking if a shortcut path can be setup every time a packet is sent over the routed path.

Some of the features of this state are:

1. Cache lookups must be fast as they are done on every packet.
2. The cache lookup must be on the destination IP address instead of the next-hop router IP address.
3. Both ACK and NAK data should be cached for the length of the holding time parameter in the NHRP response.

Since state must be maintained, the questions of where to maintain it, how to manually managed it, and how to selectively override it need to be addressed. No matter where this state information is kept, a method for manually examining and changing this state information must be provided. This is essential to insure that the network is operating properly.

There are several possible locations for storing this state information, they are:

1. Store state in the 'ARP' table. This is the traditional location for this IP to ATM address mappings. This table must be extended to handle the caching of negative (routed path) information. This solution provides a system wide service that may be used by the NHC.
2. Store state in the IP routing table. This is the traditional location for the local/remote state information.

3. Store state in an ATM MIB structure. This is the traditional location for storing ATM VCC data. It also provides a system wide service that is geared toward ATM services. This avoids munging the 'ARP' table to hold negative data.
4. Store state in the TCP Process Control Block. This allows a per process tailoring of shortcut or routed path information. This works well for TCP connections, but not UDP style services.
5. Store state in the socket structure. This also allows per process tailoring of the state information.
6. Store state in a newly defined table.

The NHC should also support both local (per-process) and global (per-system) state. This would allow a system wide default while allowing a specific application to tailor the operation for a specific task. For example assume a site runs both a DNS server and FTP server on a single host. Inter LIS communications to the DNS server should take the routed path to avoid setup overhead. While an FTP session would benefit from the shortcut path to improve performance. Supporting both operations from a single client will require both a global state (e.g. use shortcut for FTP) and a local state (e.g. use routed path for DNS).

5.1 Using TCP

TCP is a connection orientated protocol that provides per-process state information using a TCP Protocol Control Block (PCB). This PCB can be used to save the shortcut/routed path state information. Using a quad-state flag that shows the USE_SHORT_CUT, TRY_SHORT_CUT, USE_ROUTED_PATH, or TRY_ROUTED_PATH states would allow each process to use the service it chooses. The advantage of this approach is that it allows per flow control over the use of the shortcut or routed path. The disadvantage is that this PCB is only created for TCP connections. UDP connections will only use the system default action.

A second option is to store this information in the socket PCB and use the socket function (setsockopt) to save this information. This option will allow both TCP and UDP applications to set a per flow action to override the system default operation. To enable this option, the IP kernel code will need to be modified to allow this quad-state flag to be set. In addition this flag will need to be checked when each packet is sent to determine the if the shortcut or routed path is being used.

5.2 Using UDP

UDP is a connectionless orientated protocol that doesn't provide any support for state information. It relies on the application to provide the necessary state information. In this case where should the state be stored? The user application could store this itself and pass this down to the kernel in some manner. Another option is to store this information in an ATM MIB structure. A third option is to allow a socket option (setsockopt) that the user application can set to override the default behavior.

5.3 Using ICMP

In keeping with the tradition of using ICMP echo packets for Internet management functions (e.g. ping, traceroute) then it will be necessary to allow these applications to run over the shortcut and routed paths. The user will need to be able to specify which path to use and a default action needs to be defined too.

6. Conclusions

NHRP provides new services and functionality for IP nodes using ATM networks. To use these services the client must store state information that describes whether a destination node is reachable via a shortcut or a routed path.

The state information should be stored on a global per-application basis with per-process override functionality. This allows short lived functions (e.g. DNS requests) and long lived requests (e.g. ftp sessions) to use different paths. Storing state only based on the destination address means that all processes must use the same path and this creates unreasonable demands on the network. To accomplish this the /etc/services file should be modified to carry a new flag to indicate the per-application default (shortcut vs. routed path) behavior.

This state information is required to avoid having the client make a call to the NHS for every packet it sends along the routed path. It is recommended that the IP routing table be modified to support a new flag. This flag will indicate whether the NHS returned an ACK or NAK to the NHRP request.

In addition, application programmers and system administrators require the ability to explicitly request a specific service (e.g. use the routed path or shortcut path). This includes the ability to verify network operation by specifying how ICMP echo requests (e.g. ping, traceroute) are handled. The NHC must support the manual setting of this state information. A new socket option that allows

the user to specify the operation needs to be supported.

To support this capability a new socket option will be created to allow the user application to control the operation of a particular connection (flow). This option will allow the user to specify that a connection use one of the following:

* USE_SYSTEM_DEFAULT. Use the shortcut or routed path based on the system configuration information for this application. (This is the default behavior.)
* USE_SHORT_CUT. If a shortcut path exists, then use it to deliver the data. If it doesn't exist, then try and create it. If the shortcut cannot be created, fail the connection and notify the user.
* TRY_SHORT_CUT. If a shortcut path exists, then use it to deliver the data. If it doesn't exist, then try and create it. If the shortcut cannot be created, try using the routed path.
* USE_ROUTED_PATH. Use the routed path regardless of whether a shortcut exists or not.
* TRY_ROUTED_PATH. If a shortcut doesn't exist, don't try and create it, use the routed path instead.

7. Security

The security issues for NHRP are addressed in other NHRP documents [2,3]. Some specific security issues for the NHC developer are discussed below.

* Address spoofing at the IP or ATM layer may allow an attacker to hi-jack an IP connection or service. This threat may be reduced by limiting the scope of the ATM routing domain. In this way only trusted IP hosts will be able to reach and use the services of the NHS.
* Denial of service attacks may be launched at both the IP and ATM layers of the NHS. At the ATM layer, the attacker may repeatedly generate signaling messages that consuming system resources thus preventing NHCs from using the NHS services. At the IP layer, the attacker may register false IP to ATM mappings thus preventing a NHC from registering the correct IP to ATM mapping.
* When a NHC creates or accepts a short-cut path it bypasses the site border router. Therefore, any security features in the border router are also bypassed. This threat may be reduced by limiting the scope of the ATM routing domain, increasing

security features in the NHC host, allowing the NHS to evaluate security features when short-cut paths are requested or a compination of all of these methods.

8. Authors' Addresses

Richard Carlson
Argonne National Laboratory

EMail: RACarlson@anl.gov

Linda Winkler
Argonne National Laboratory

EMail: lwinkler@anl.gov

9. References:

[1] Laubach, M. and J. Halpern, "Classical IP and ARP over ATM", RFC 2225, April 1998.

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

[3] Cansever, D., "NHRP Protocol Applicability Statement", RFC 2333, April 1998.

[4] Luciani, J., "Classical IP to NHRP Transition", RFC 2336, July 1998.

[5] Rekhter, Y. and D. Kandlur, "Local/Remote Forwarding Decision in Switched Data link Subnetworks", RFC 1937, May 1996.

10. Full Copyright Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.