

Network Working Group
INTERNET DRAFT
Expires September 1998

Turner S., Staegemeir B.
SITA / Equant R&D
March 1998

Differentiated Services over Symmetric NHRP Shortcuts
<[draft-turner-diff-nhrp-00.txt](#)>

Status of this Memo

This document is an Internet-Draft. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its Areas, and its Working Groups. Note that other groups may also distribute working documents as Internet Drafts.

Internet Drafts are draft documents valid for a maximum of six months. Internet Drafts may be updated, replaced, or obsoleted by other documents at any time. It is not appropriate to use Internet Drafts as reference material or to cite them other than as a "working draft" or "work in progress".

Please check the `l1d-abstracts.txt` listing contained in the internet-drafts Shadow Directories on `nic.ddn.mil`, `nisc.nsf.net`, `nic.nordu.net`, `ftp.nisc.sri.com`, or `munari.oz.au` to learn the current status of any Internet Draft.

Abstract

There is a need, for relatively simple and scaleable methods of providing differentiated classes of service for IP traffic over various WAN media, to support different types of applications and specific business requirements. The Differentiated Services approach to providing quality of service (QoS) in Networks, employs a small, well-defined set of building blocks from which a variety of services may be built. A small bit-pattern in each packet in the IPv4 "TOS octet" or in the IPv6 "Traffic Class octet" is used to mark a packet to receive a differentiated forwarding treatment or per-hop behavior at each network node. IETF Working Groups will standardize a common layout for both octets, called the "DS byte" superseding the IPv4 TOS octet definitions or [RFC 1349](#) [1].

The goal of this contribution is to suggest how the Next Hop Resolution Protocol (NHRP) may be extended and deployed in conjunction with the IP ToS octet (DS-Byte) to create both differentiated forwarding and differentiated class of service over an NBMA network such as ATM or Frame-Relay.

In the proposed solution, any SVC created between the same two points in an NBMA network will be used bi-directionally by traffic with the same class of service. This is achieved by extending the scope of the existing NHRP protocol to include QoS information in the extensions part of both the NHRP resolution request and resolution reply, and by implementing a NBMA addressing scheme that dynamically maps CoS, to subaddresses that are unique within the NBMA network.

The proposal is compliant with the NHRP specifications, and every effort is made by the proposal to build on the work currently underway within the IETF that is attempting to standardize on the use of the ToS octet (DS-byte) within the IP datagram.

1. NHRP Default Operation

Within the scope of existing definitions [2], NHRP shortcuts may be created between hosts, or ingress and egress routers at the edges of an NBMA network, with NHRP cache entries mapping remote layer 3 prefixes (IP hosts, subnets or BGP next-hops) onto SVCs. These cache entries may only be added after specific datagrams matching defined trigger criteria send an NHRP resolution request, and when that NHRP request is successfully returned. If in the NHRP resolution reply, the remote NBMA subaddress of the NHS maps to an existing SVC that directly interconnects NHC and NHS, then the best match prefix in the forwarding table, that allowed the initial forwarding of the datagram over the default path, is typically added to the NHRP cache, and is mapped onto the existing SVC. This means that all traffic flows between the same pair of ingress and egress points whose L3 destination address is mapped by the NHRP cache, will be merged onto the same NHRP shortcut.

This default behaviour is very useful, particularly for network bootstrapping and for traffic engineering purposes, in order to minimize the number of routed hops and to reduce congestion in the core of the IP network.

2. Extension of NHRP for Differentiated Forwarding

The current applicability of NHRP severely limits the possibilities to use NHRP as a mechanism, or as part of a mechanism, to create Differentiated Classes-of-Service (CoS) across an NBMA network. This is because the NHRP cache entries that indicate whether a shortcut should be used or not, are based on layer 3 prefixes, and although a delay sensitive application may in fact trigger the shortcut SVC, once created, this shortcut will be used by all traffic whose L3 destination is included in the cache. This in turn means that all traffic with a mixture of classes of service will be competing for

resources over the same forwarding path, unless a mechanism is found which allows multiple shortcuts, and where path selection is based on the class of service information in the traffic flows.

However, using multiple SVCs between the same ingress and egress pair is problematic for two main reasons :

- i) You need to introduce an additional element to the trigger criteria that specifies whether to use an existing SVC or not.
- ii) If another SVC is created for a specific policy at one end, in order to use this SVC in full duplex mode, both the policy and the mapping to an SVC must be known and synchronized at each end.

Problem (i) is not insurmountable, but the corresponding changes to the NHRP cache structure may make it difficult to implement rapid packet forwarding over the correct SVC.

Problem (ii) is more difficult, and requires the introduction of either a "central policy control point" ("one more server") and out-of-band synchronisation, or using a new "protocol" that allows end-to-end in-band correlation of layer 3 information with SVCs. This is definitely a non-trivial problem!

A compromise is suggested to permit multiple SVCs between the same ingress and egress points, and to allow both differentiated forwarding and differentiated class of service.

The primary purpose of this strategy is to allow the creation of shortcuts, and through the use of differentiated forwarding signaled by the ToS byte in IP datagrams, to prevent "low class" traffic mixing with "high class" traffic. Thus we create more performance determinacy.

3. Per Class-of-Service Shortcuts

The proposal has three components:

- i) A simple way to map an IP packet with a specific Class of Service information in the DS-byte to a corresponding SVC with an appropriate layer 2 traffic class. The number of possible SVCs between the same source and destination prefix is therefore limited to the number of relevant DS-byte encodings, e.g. as defined in [RFC 1349](#). This does not imply that each defined CoS always maps onto a unique SVC, since different CoS may be grouped and mapped onto a shared SVC.
- ii) The local association or mapping by Next Hop Resolution Clients (NHC) and Next Hop Resolution Servers (NHS), between a single class of service or a group of classes of service, and a unique NBMA subaddress.

- iii) A protocol extension to ensure that all SVC based shortcuts are used in both directions. This avoids the protocol and processing overhead for SVC creation in the reverse direction, and eliminates any additional set up delay and potential performance problems caused by asymmetric paths.

The proposed "CoS enabled NHRP" works in the following way :

- i) An NHRP request policy is directly associated with the IP ToS byte or specific elements in it. An obvious example would be triggers based on the four precedence bits and its eight defined classes of services in [RFC 1349](#).
- ii) This policy should be consistently defined through NHRP "triggers" on all NHCs in the NBMA network.
- iii) Each NHC and NHS makes a local association between a Class of Service, identified by the ToS byte, and a globally unique NBMA subaddress. This NBMA subaddress may be formed by appending or including a locally managed addressing element termed the "CoS designator", to the globally known and unique NBMA subaddress. e.g. In the case of a given router or host connected to an ATM network, each CoS is mapped to a specific value of selector byte that is then appended to the same 13 byte prefix and 6 byte ESI, to form a set of unique NSAPs. Each NSAP thus globally identifies the NHS or NHC, and provides local significance as to the CoS assigned to each SVC shortcut. This association should be dynamically mapped according to the local implementation and means that the number of NBMA addresses or subaddresses per NHRP station must be greater than or equal to the number of CoS available across the NBMA subnetwork.
- iv) When an NHRP client forwards a datagram into the NBMA cloud with a ToS byte that matches the trigger policy, an NHRP resolution request is created and sent. (It is assumed that no SVC shortcuts exist). This NHRP request contains the ToS byte in the IP packet that triggered it, and is copied into the Extensions part of the NHRP resolution request.
- v) When the NHRP request is received by the NHS at the egress point of the NBMA network, the ToS extension is examined, and the NBMA address returned in the reply is specific to that ToS. i.e. the CoS designator that is locally associated with the value in the ToS extension is included in the NBMA address returned in the NHRP resolution reply.

- vi) When the NHRP reply is received by the NHC that triggered the initial request, an NHRP cache entry is created that maps the L3 reachability information associated with the destination to the NBMA address contained in the resolution reply. This NHRP cache entry is now extended to include the ToS information sent in the resolution request and subsequently returned in the resolution reply. An SVC to the remote NHS for this CoS can be created immediately or on receipt of a subsequent packet that matches the cached information. The time at which the SVC is created and any L2 QoS related signaling parameters are implementation specific.)
- vii) Subsequent packets sent by the NHC are now compared to the NHRP cache entry in terms of destination IP address and the ToS byte.

Now we distinguish three cases :

- i) If a match is found on both conditions, then the packet is forwarded over the associated SVC.
- ii) If a match is not found, and the QoS information in the packet header matches the trigger policy, then a new NHRP request is sent out as described previously. When the reply is received and both destination NBMA address and ToS information is the same as an existing NBMA address/ToS pair, then the destination IP prefix is added to the NHRP cache, and the entry is mapped to the existing SVC.
- iii) In the case of (ii), if the reply contains an NBMA address /ToS pair that does not map to an existing NHRP cache entry, then a new SVC is requested to the NBMA subaddress contained in the reply, and a new cache entry created as before.
- viii) In the reverse path, i.e. when the roles of NHC and NHS are reversed, an identical process is followed. In this case, since an NHS should consistently return an NBMA subaddress that maps to a given CoS or group of CoS, then it is straightforward to determine if an SVC with the given CoS already exists, even if the SVC was remotely initiated.

4. Full Duplex Mode of Operation and Asymmetric shortcut avoidance

To avoid two SVCs with the same CoS being used unidirectionally between the same pair of NHRP stations, the following simple algorithm is proposed.

When an NHRP resolution request is received by the NHS, a lookup is performed that keys on the source NBMA subaddress without CoS

designator i.e. without the selector in the ATM case.

If no match is found, then continue as described previously.

If a match is found, extract the requested CoS in the extensions part of the NHRP Resolution request and:

- i) If neither the CoS nor the CoS designator from the NHRP resolution request are found in the CoS mapping table, then continue.
- ii) If both the CoS and the CoS designator are found in the table, and they are mapped together, then continue.
- iii) If the CoS is found but it maps to a different CoS designator, then send a Resolution Reply with a NAK code of 4 "Administratively Prohibited"
- iv) If the CoS designator is found but it maps to a different CoS, then send a Resolution Reply with a NAK code of 4, "Administratively Prohibited"

When a resolution reply is received by the NHC, then do the following:

- i) If NAK is received, then abort all attempts to include L3 destination in NHRP cache.
- ii) If the reply is valid, but the Compulsory bit is clear, assume CoS 0, i.e. shortcut forwarding over single SVC only to NHS. (see below)
- iii) If the reply is valid, and the Compulsory bit is set, then:
 - a) Perform a lookup based on the source NBMA subaddress without CoS designator i.e. without the selector in the ATM case.
 - b) If neither the CoS nor the CoS designator from the NHRP resolution request are found in the CoS mapping table, then continue.
 - c) If both the CoS and the CoS designator are found in the table, and they are mapped together, then continue.
 - d) If the CoS is found but it maps to a different CoS designator, then abort all attempts to include L3 destination in NHRP cache.
 - e) If the CoS designator is found but it maps to a different CoS, then abort all attempts to include L3 destination in NHRP cache.

5. Backwards compatibility with existing NHRP Implementations

The "Compulsory" bit should be left clear in the NHRP resolution request to allow backwards compatibility with existing NHSSs. If the egress router understands the extension and acts upon it, i.e. is performing CoS mapping, then this C bit should be set to 1 in the reply. This will allow the requestor to know whether to perform the asymmetric checking or not, and whether to create multiple SVCs to the same NHS.

6. Issues and Scalability

The techniques described above would allow the creation of a strongly differentiated class of service within an IP network that primarily uses an NBMA network for interconnection between hosts and or routers. The shortcuts created by NHRP will be exclusive per associated Class-of-Service (rather than per flow), and will be used in full-duplex mode. The forwarding overhead to use a shortcut is "simply" using a combination of both destination address and QoS to hash the NHRP cache. The primary limitation could be the availability of some form of "sub-addressing" scheme for the NHRP clients and servers that allows the local association between the called and calling addresses and a CoS.

This solution scales well due to a simple but consistent forwarding Policy definition, the limited number of SVCs, and since the number of NHRP cache entries is limited per CoS per destination prefix.

7. Security Considerations

Security is not addressed in the current version of this document

References

- [1] Almquist, P. "Type of Service in the Internet Protocol Suite".
[RFC 1349](#), July 1992.
- [2] NMBA Next Hop Resolution Protocol (NHRP) , J. Luciani,
D. Katz, D. Piscitello and B. Cole. Work in progress.

Authors' Address

Steve Turner
SITA Network Projects Department.
1041, Route des Dolines
F-06560 Valbonne / France
Phone +33 4 92 96 63 17
email st@pt.nce.sita.int

Bernd Staegemeir
SITA Research & Development
1041, Route des Dolines
F-06560 Valbonne / France
Phone +33 4 92 96 65 40
email bst@ed.nce.sita.int

