Network Working Group
Internet-Draft
Expires: November 13, 2003

C. de Launois
O. Bonaventure
UCL/INGI
May 15, 2003

NAROS : Host-Centric IPv6 Multihoming with Traffic Engineering
draft-de-launois-multi6-naros-00.txt

Status of this Memo

Copyright Notice

Abstract

More and more ISPs are multihomed and need to engineer their interdomain traffic. Unfortunately, both multihoming and traffic engineering contribute to the growth of the BGP routing tables. For IPv6, a better solution for multihoming and traffic engineering is required. This document proposes a host-centric IPv6 multihoming solution that relies on the utilization of a "Name, Address and ROute System" (NAROS) server. A key advantage of using this server is that it allows a multihomed site to engineer its interdomain traffic without transmitting any BGP message.

Table of Contents

1. Introduction

The size of BGP routing tables in the Internet has been growing dramatically during the last years. Their size has risen from 70,000 entries in january 2000 to about 140,000 at the beginning of 2003 [19]. The current size of those tables creates operational issues for some Internet Service Providers and several experts [1] are concerned about the increasing risk of instability of BGP.

Part of the growth of the BGP routing tables is due to the fact that, for economical and technical reasons, many ISPs and corporate networks wish to be connected via at least two providers to the Internet. Nowadays, at least 60% of those networks domains are connected to two or more providers [2]. Other factors include address fragmentation or failure to aggregate [3]. Once multihomed, a domain will usually want to engineer its interdomain traffic to reduce its Internet bill. Unfortunately, the available interdomain traffic engineering techniques [4] are currently based on the manipulation of BGP attributes which contributes to the growth and the instability of the BGP routing tables.

Although several solutions to the IPv6 multihoming problem have been discussed within the multi6 working group of IETF, few have addressed the need for interdomain traffic engineering.

The NAROS approach presented in this document is a host-centric IPv6 multihoming solution which allows sites to engineer their incoming and outgoing interdomain traffic without any manipulation of BGP messages. It relies on the utilization of several IPv6 addresses per host, one from each provider. The basic principle of NAROS is that before transmitting packets, hosts contact the NAROS service to determine which IPv6 source address they should use to reach a given destination.

In the second section, we briefly present the technical and

economical reasons for multihoming in the Internet. In the third section, we describe the NAROS architecture and explain how it supports multihoming and traffic engineering. In the fourth section, the NAROS messages formats are detailed. Finally, security considerations are discussed in the fifth section.

2. Multihoming Issues

IPv6 multihoming solutions are significantly different from IPv4 ones because they must allow the routing system to better scale. Morever, the IPv6 address space is much larger, which gives more freedom when designing multihoming. An IPv6 host may have several global addresses. Paradoxically this can help in reducing the BGP table

sizes but it requires that hosts correctly handle multiple addresses. Requirements for IPv6 multihoming are stronger and multiple [5]. In this document, we essentially focus on the following requirements.

2.1 Fault Tolerance

Sites connect to several providers mainly to get fault tolerance. A multihoming solution should be able to insulate the site from both link and ISP failures.

2.2 Route Aggregation

Every IPv6 multihoming solution is required to allow route aggregation at the level of their providers [1], [5]. This is essential for the scalability of the interdomain routing system.

2.3 Source Address Selection

A multihomed IPv6 host may have several addresses, assigned by different providers. When selecting the source address of a packet to be sent, a host could in theory pick any of these addresses. However, for security reasons, most providers refuse to convey packets with source addresses outside their address range [6]. So, the source address selected by a host also determines the upstream provider used to convey the packet. This has a direct impact on traffic engineering. Moreover, if a host selects a source address belonging to a failed provider, the packet will never reach its destination. Thus, a mechanism must be used to select the most

appropriate source address.

2.4 Destination Address Selection

When a host in the Internet contacts a multihomed host, it must determine which destination address to use. The destination address also determines the provider used. If a provider of the multihomed site is not available, the corresponding destination address cannot be used to reach the host. So we must make sure that an appropriate destination address is always selected.

2.5 Traffic Engineering

A multihomed site should be able to control the amount of inbound and outbound traffic exchanged with its providers. It should also be able to prefer one provider over others for some routes.

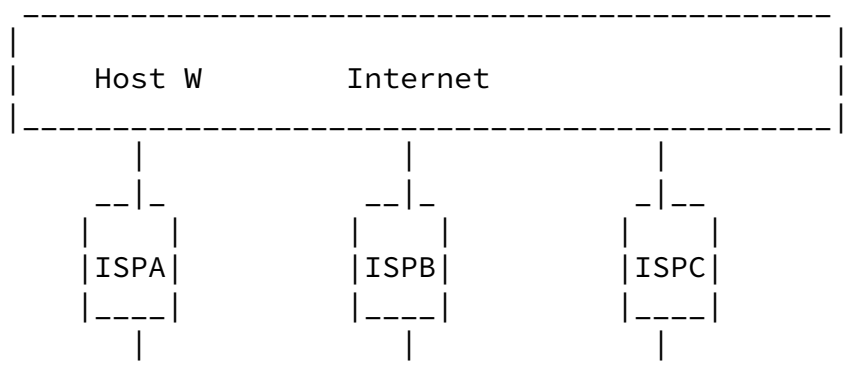
2.6 ISP Independence

It is desirable that a multihoming solution can be set up

independently without requiring cooperation of the providers.

3. The NAROS Service

Figure 1 illustrates a standard multihomed site. Suppose three Internet Service Providers (ISPA, ISPB and ISPC) provide connectivity to the multihomed site. The site exit router connecting with ISPA (resp. ISPB and ISPC) is RA (resp. RB and RC). Each ISP assigns a site prefix to the multihomed site. The prefixes are then used to derive one IPv6 address per provider for each host interface.



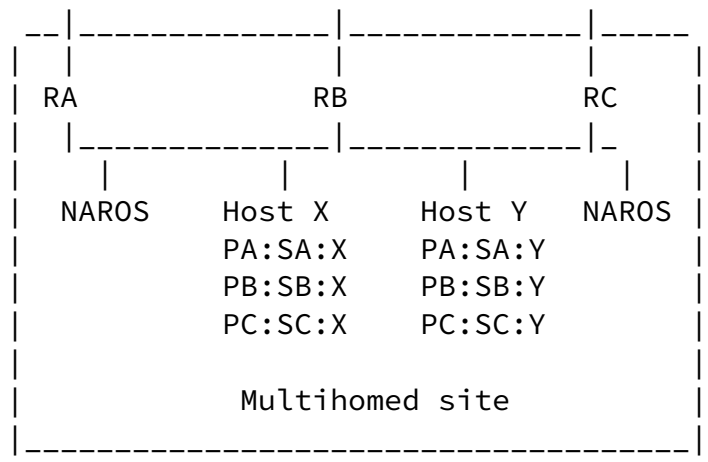


Figure 1

In the NAROS architecture, the site sends packets with ISPA addresses only to ISPA, and ISPA only announces its own IPv6 aggregate to the global Internet.

Since each host has several IPv6 addresses, it must decide which address to use when transmitting packets. The basic principle of our solution is to let the NAROS service manage the selection of the source addresses. This address selection will influence how the traffic flows through the upstream providers and a good selection method will allow the site to engineer its interdomain traffic.

We now consider in details how the NAROS service addresses four main

issues : source and destination address selection, fault-tolerance and traffic engineering.

3.1 Source Address Selection

When a host initiates a connection with a correspondent, it must determine the best source address to use among its available addresses. The source address selection algorithm described in [7] already provides a way to select an appropriate address. However, this selection is arbitrary when a host has several global-scope IPv6 addresses as in the host-centric multihoming case.

The principle that we propose is that the host asks the NAROS service which source address to use. It complements in this way the default

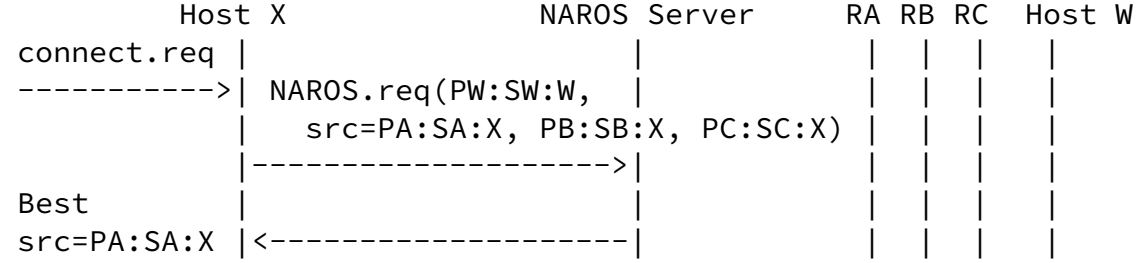
IPv6 source address selection algorithm [7].

In its simplest form, the basic NAROS service is independent from any other service. A NAROS server does not maintain state about the internal hosts. It is thus possible to deploy several NAROS servers in anycast mode inside a site for redundancy or load-balancing reasons. A NAROS server can also be installed on routers such as the site exit routers. The NAROS protocol can run over UDP, as it is described in this document. It can also run over another protocol like ICMPv6 [8]. The NAROS protocol contains only two messages : NAROS request and NAROS response.

The first message is used by a client to request which source address to use to reach a given destination. The parameters included in a NAROS request are at least the destination address of the correspondent and the source addresses currently allocated to the client. The NAROS server should only be contacted when the default source address selection procedure [7] cannot select the source address.

The NAROS response message is sent by a NAROS server and contains the source address to be used by the client. The parameters include at least the selected best source address, a prefix and a lifetime. It tells that the client can use the selected source address to contact any destination address matching the prefix. These parameters remain valid and can be cached by the client during the announced lifetime.

Each host knows from its configuration the address of the NAROS server. This address can be an anycast address.



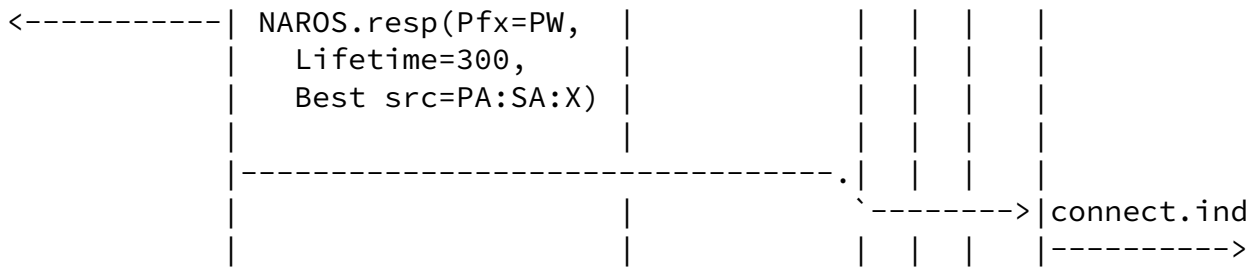


Figure 2 shows an example of how the NAROS messages and parameters are used. The exact format of the NAROS message is described in Section 4.2. When Host X sends its first packet to remote Host W (PW:SW:W), it issues a NAROS request in order to obtain the source address to use to reach Host W. Upon receipt of the request, the NAROS server identifies the prefix PW associated with Host W and selects for example PA:SA:X as the best source address. The prefix can be determined arbitrarily, e.g. using the /48 prefix corresponding to the destination address. Another solution is to extract from a BGP table the prefix associated with the destination. The server then indicates the lifetime (e.g. 300 seconds) of these parameters in the NAROS response message.

After having processed the reply, Host X knows that it can use PA:SA:X to reach any destination inside prefix PW, including Host W. The selected source address should be used for the whole duration of the flow, in order to preserve the connection. If new TCP connections or UDP flows to the same destination are initiated before the announced lifetime expires, the client can use the cached parameter. Otherwise the host must issue a new NAROS request and it may get a different source address for the same destination. By using appropriate values for the lifetime and the prefix in the NAROS response, it is possible to reduce the number of NAROS requests sent by hosts.

A simulation of the influence of various NAROS parameters and the number of messages exchanged may be found in [9].

3.2 Destination Address Selection

A second case is when Host W on the Internet needs to contact Host X

in the multihomed site. It first issues a DNS request. The DNS server of the multihomed site could reply with all the addresses associated to Host X. At worst, Host W will try the proposed addresses one by one. Eventually, a connection will work. A better solution to this problem would be to integrate an extended NAROS service with the DNS server to choose the address to be returned. This choice could depend on the source of the DNS request, a policy defined by the administrator, or the traffic engineering requirements.

3.3 Fault Tolerance

A third problem to consider is what happens when one of the upstream providers fails. As in the solution described in [10], [11], the site exit routers use router advertisement messages to communicate to hosts the available prefixes [12], [13]. When a provider crashes, the site exit router connected to this provider detects the event and advertises a null preferred lifetime for that prefix. A client can take this event into account by immediately asking new parameters to the NAROS server. More generally, a host can ask updated parameters each time it detects a failure which affects one of its communications. Once the new source address is known, HIP [14], SCTP [15], IP mobility [16] or other mechanisms [17] can be used in order to preserve the established TCP connections in case of ISP failure.

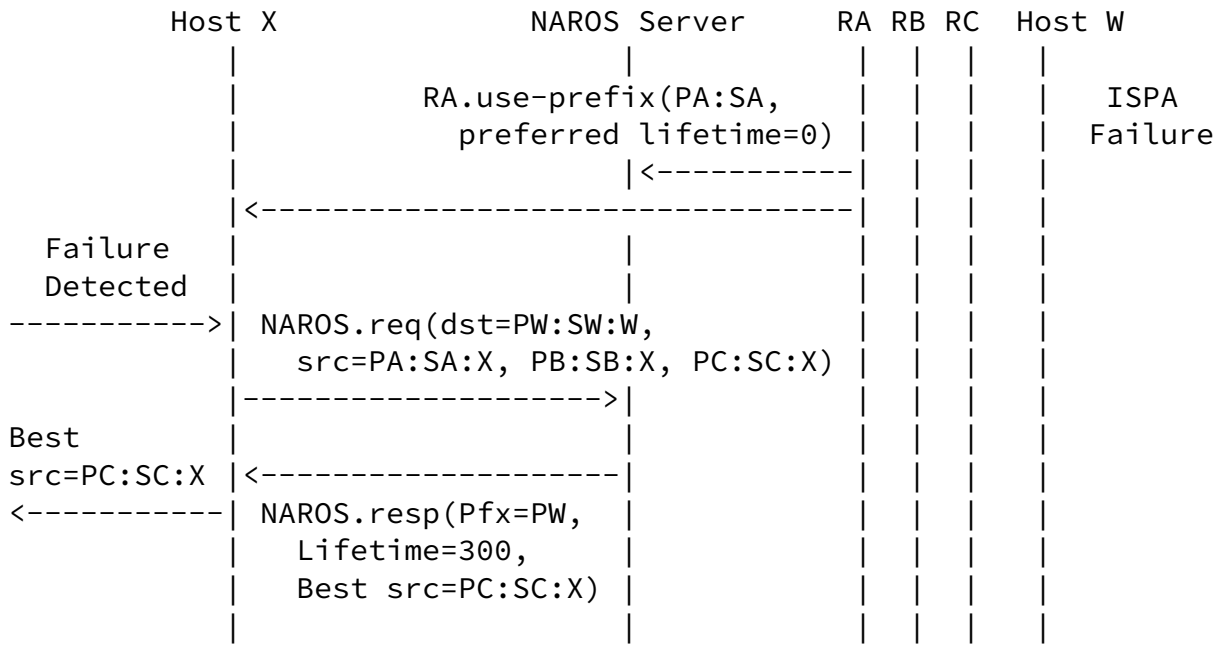


Figure 3

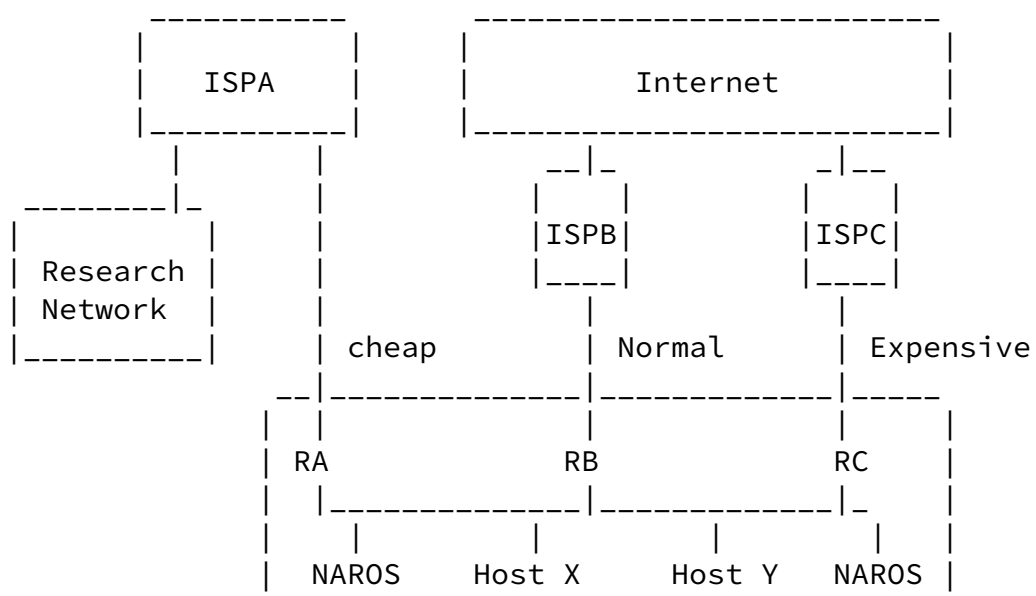
In Figure 3, consider for example that ISPA becomes unavailable. The site exit router connected to ISPA detects the failure and advertises a null preferred lifetime for prefix PA. The NAROS server

immediately takes this advertisement into account and future NAROS replies will not contain this prefix. Host X will also receive this advertisement. The standard effect is that it should no longer use this source address for new TCP or UDP flows. If Host X is currently using a deprecated address, it can issue a new NAROS request to choose among its other available source addresses. The host can then use IP mobility mechanisms to switch to the new source address in order to maintain its connection alive.

3.4 Interdomain Traffic Engineering

When a host selects a source address, it also selects the provider through which the packets will be sent. Since the source address to use is selected by the NAROS service, this can naturally be used to perform traffic engineering.

For example, in order to equally balance the traffic among the three providers in Figure 1, a NAROS server can use a round-robin approach. Each time it receives a NAROS request, the server selects another provider and replies with the corresponding source address. Except when a provider fails, this source address, and thus the upstream provider, remains the same for the whole duration of the flow. Note that this solution allows traffic engineering without injecting any information in the internet routing system. Moreover, the NAROS service can easily support unequal load distribution, without any additionnal complexity.



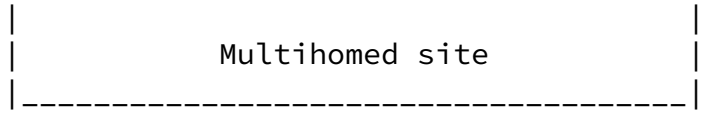


Figure 4

Another example is illustrated in Figure 4. ISPA only provides connectivity between the multihomed site and a research network. The link to ISPA is cheap while the link to ISPC is expensive. In this example, a NAROS server can select ISPA for the research traffic, and balance the Internet traffic between ISPB and ISPC with a higher preference for ISPB.

Beside the above functionalities, the NAROS approach has several advantages. First, no coordination is needed between the providers of the multihomed site. A provider only delegates a prefix to the site. This makes the solution applicable even for small sites which cannot afford provider-independent addresses. Moreover, the multihomed site is able to engineer its traffic without any agreement with its providers. Finally, since routes to addresses delegated by one provider are not announced to other providers, full route aggregation is possible, at the price of multiple addresses per hosts.

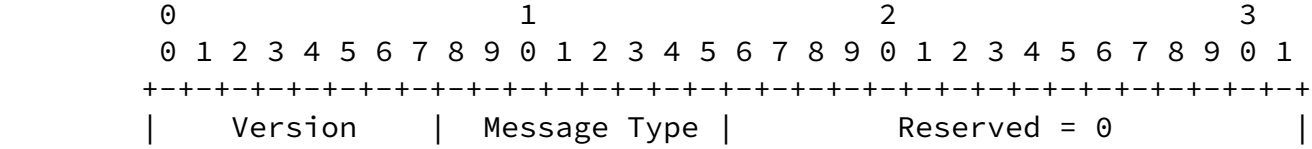
4. The NAROS protocol

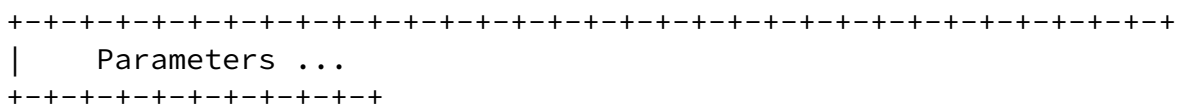
4.1 Specification of Requirements

The keywords "MUST", "MUST NOT", "REQUIRED", "SHOULD", "SHOULD NOT", "SHALL", "SHALL NOT", "MAY" and "MAY NOT" that appear in this section of this document are to be interpreted as described in [18].

4.2 Message Types

NAROS messages consist of two mandatory fields, Version and Message Type, followed by one or more required parameters. Message format is shown below :





The version field is a single byte that specifies the NAROS version number that is being used. The current NAROS version number is 1.

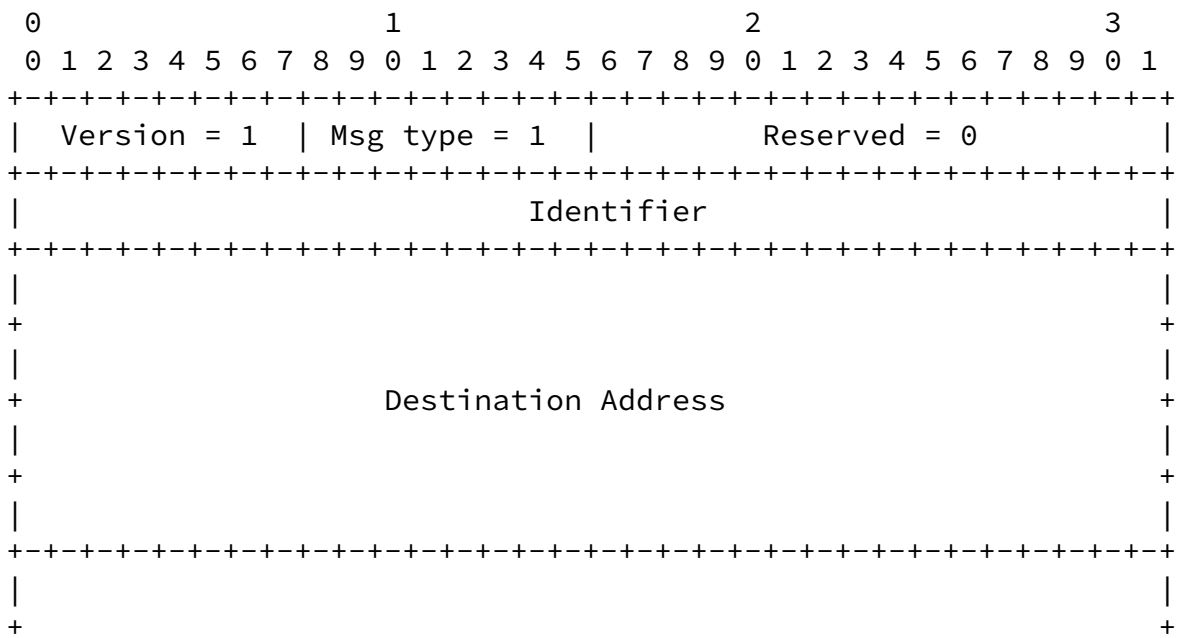
The message type field is a single byte and specifies the message contained in the current packet. A NAROS message MUST NOT be longer than 4096 bytes and there may be only one message per packet.

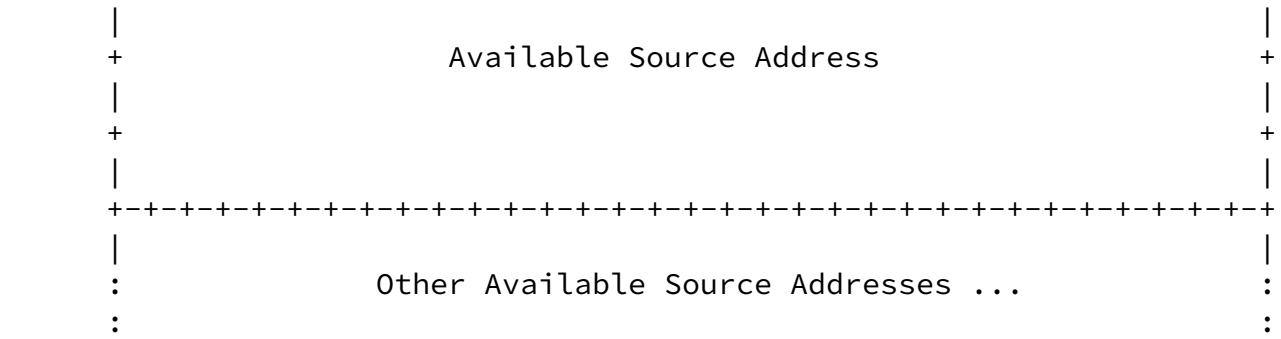
Defined message types are :

Value	Message	
1	NAROS_REQUEST	Host -> Server
2	NAROS_RESPONSE	Server -> Host
3	NAROS_ERROR_RESPONSE	Server -> Host

4.2.1 NAROS_REQUEST

A NAROS_REQUEST is sent by a NAROS client to request its connection parameters. The format of the message is :





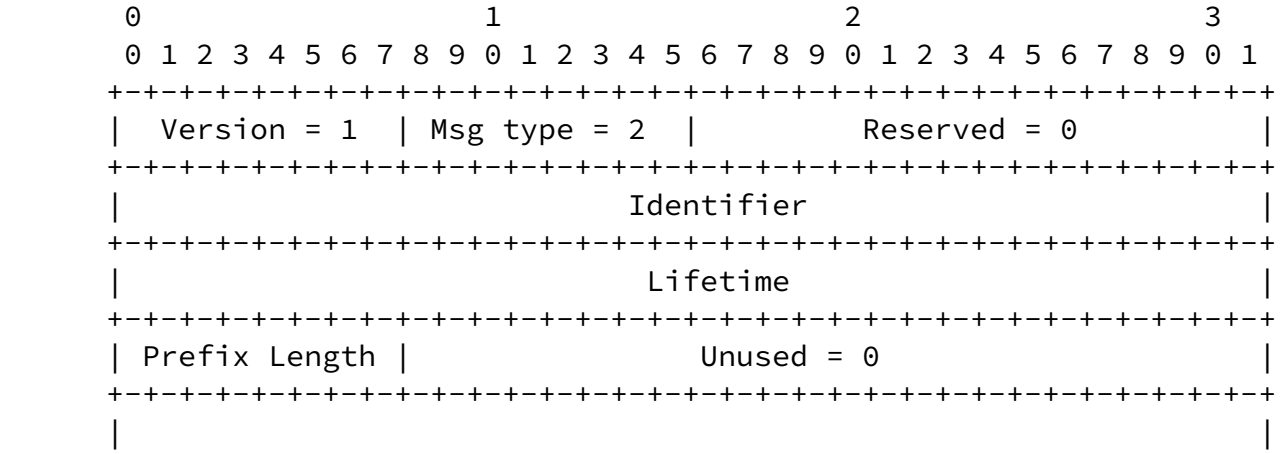
The Identifier field uniquely identifies the NAROS request. The Destination Address field is the address of the destination host that the client wants to contact. The Available Source Address fields are the source addresses currently allocated to the client.

When selecting a source address, the client MUST first use the source address selection algorithm as defined in [7]. However, the 8th rule, i.e. use longest matching prefix, MUST be superseded by the NAROS source address selection procedure. A NAROS_REQUEST message

MUST be sent by the client when it initiates a communication with a remote host for the first time. The client MUST specify in the message all the source addresses it is able to use to reach the destination, i.e. its global-scope non-deprecated IPv6 addresses.

4.2.2 NAROS_RESPONSE

The NAROS_RESPONSE message is sent by a NAROS server and specifies which is the best source address to contact the destination.



The Identifier field specifies the request for which the NAROS_RESPONSE message is a reply. It must be the same identifier as the identifier contained in the associated request.

The Lifetime field defines how long the information contained in this message can be cached by the client.

The Prefix Length and Destination Prefix fields defines the destination prefix for which the best source address contained in this message applies. The Prefix Length field indicates the length

in bits of the destination IP address prefix. A length of zero indicates a prefix that matches all IP addresses. The Destination Prefix field contains the destination IP address prefix followed by enough trailing bits to fill the field. Note that the value of trailing bits is irrelevant.

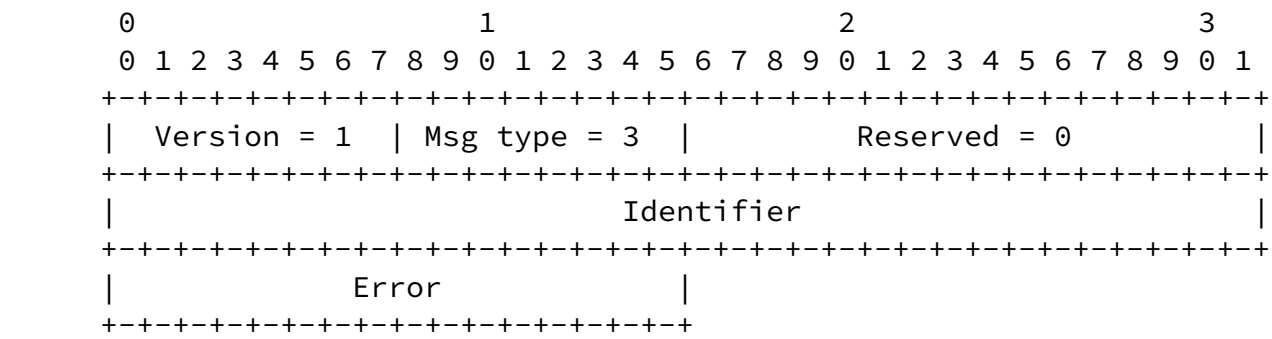
The Best Source Address field specifies the source IP address that the server determined to be the best to reach the destination prefix.

When receiving a NAROS_RESPONSE message, the client MUST first check the validity of the selected best source address, i.e. it must check that the source address is available and not deprecated. If the client determines that the best source address is not a valid choice, then it MAY use another available source addresses. Otherwise it SHOULD use the selected source address for all communications towards

any destination belonging to the destination prefix, until the lifetime expires. The selected source address for the destination prefix SHOULD be cached during the specified lifetime.

4.2.3 NAROS_ERROR_RESPONSE

A NAROS_ERROR_RESPONSE is used to provide error messages from a NAROS server to a NAROS client. Multiple errors MAY NOT be reported in the same NAROS_ERROR_RESPONSE. In situations where more than one error has occurred, the NAROS server MUST choose only one error to report. The format of the message is :



The Identifier field specifies the request for which the NAROS_ERROR_RESPONSE message is a reply. An NAROS_ERROR_RESPONSE message MUST only be transmitted by a NAROS server, and only in response to a request from a NAROS client. A NAROS client that detects an error in a message received from a NAROS server MUST silently discard the message.

The values for the Error field are defined in Section 7.

5. Security Considerations

A NAROS server should take all measures deemed necessary to prevent its clients from performing intentional or unintentional denial-of-service attacks by issuing a large number of requests. Moreover, NAROS messages received by a client should be authenticated since the introduction of malicious NAROS_RESPONSE messages could divert traffic from its regular path.

However, the NAROS service allows all the providers of the multihomed

site to perform ingress filtering, which benefits to security.

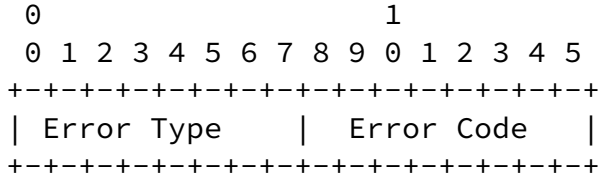
6. Conclusion

With the solution proposed in this document, several provider-aggregatable IPv6 addresses are allocated to each host inside a multihomed site. When a host needs to communicate with a destination outside its site, it contacts its NAROS server to determine the appropriate source IPv6 address to be used. The NAROS server does not maintain any per-host state and could easily be deployed as an anycast service inside each site. An advantage of the utilization of the NAROS server is that by selecting the addresses to be used by the hosts, the NAROS server influences the flow of traffic with its providers. This allows the NAROS server to indirectly, but efficiently, engineer its interdomain traffic, without manipulating any BGP attribute. Moreover, full route aggregation is maintained and the NAROS service can be set up independently from the providers. Changes are limited to hosts inside the multihomed site. Legacy hosts are still able to work, even if they cannot benefit from site-multihoming.

Performance evaluations of the NAROS protocol can be found in [9].

7. Appendix A : NAROS Error Numbers

This section provides descriptions for the error values in the NAROS_ERROR_RESPONSE message. The 2-byte length Error field is divided into a 1-byte Error Type and a 1-byte Error Code.



The currently defined error values are shown below.

Type	Type	Code	Error Name

1	General errors	1	1	UNKNOWN_ERROR
2	Message errors	1	2	INTERNAL_SERVER_ERROR
3	Routing errors	1	3	UNSUPPORTED_NAROS_VERSION
		2	1	ILLEGAL_MESSAGE
		2	2	BAD_MESSAGE
		3	1	NO_ROUTE
		3	2	ADMINISTRATIVELY_PROHIBITED

1 : General errors

UNKNOWN_ERROR. An error that cannot be identified has occurred. This error should be used when all other error messages are inappropriate.

INTERNAL_SERVER_ERROR. An NAROS server application has detected an unrecoverable error within itself.

UNSUPPORTED_NAROS_VERSION. An NAROS client sent a message with a version number that is not supported by the NAROS server.

2 : Message errors.

The server uses these errors when it detects that a message is malformed, as well as when it does not understand a message.

ILLEGAL_MESSAGE. The message contains illegal values for one or more fields.

BAD_MESSAGE. The message is malformed and server parsing failed.

3 : Routing errors.

The server uses these errors when it is unable to select the best source address needed by the client to reach a destination.

NO_ROUTE. The server has no route towards the destination.

ADMINISTRATIVELY_PROHIBITED. The server has a route towards the destination but it is administratively prohibited.

References

[1] Atkinson, R., "IAB Concerns & Recommendations Regarding Internet Research & Evolution", Internet Draft draft-iab-research-funding-00, February 2003.

[2] Agarwal, S., Chuah, C. and R. Katz, "OPCA: Robust Interdomain Policy Routing and Traffic Control", Proc. OPENARCH, 2003.

[3] Bu, T., Gao, L. and D. Towsley, "On Routing Table Growth", Proc. IEEE Global Internet Symposium, 2002.

[4] Quoitin, B., Uhlig, S., Pelsser, C., Swinnen, L. and O. Bonaventure, "Interdomain traffic engineering with BGP", IEEE Communications Magazine, May 2003.

[5] Abley, J., Black, B. and V. Gill, "Goals for IPv6 Site-Multihoming Architectures", Internet Draft draft-ietf-multi6-multihoming-requirements-04, April 2003.

[6] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address SpoofingNeighbor Discovery for IP Version 6 (IPv6)", RFC 2267, January 1998.

[7] Draves, R., "Default Address Selection for IPv6", Internet Draft draft-ietf-ipv6-default-addr-select-09, August 2002.

[8] Conta, A. and S. Deering, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address SpoofingNeighbor Discovery for IP Version 6 (IPv6)", RFC 2463, December 1998.

[9] de Launois, C., Bonaventure, O. and M. Lobelle, "The NAROS Approach for IPv6 Multi-homing with Traffic Engineering", URL http://www.info.ucl.ac.be/people/delaunoi/, Submitted QoFIS, April 2003.

[10] Dupont, F., "Multihomed routing domain issues for IPv6 aggregatable scheme", Internet Draft draft-ietf-ipngwg-multi-isp-00, September 1999.

[11] Huitema, C. and R. Draves, "Host-Centric IPv6 Multihoming", Internet Draft draft-huitema-multi6-hosts-01, June 2002.

[12] Narten, T., Nordmark, E. and W. Simpson, "Neighbor Discovery for IP Version 6 (IPv6)", RFC 2461, December 1998.

[13] Thomson, S. and T. Narten, "IPv6 Stateless Address Autoconfiguration", RFC 2462, December 1998.

[14] Moskowitz, R., "Host Identity Payload Architecture", Internet

[15] Coene, L., "Multihoming issues in the Stream Control Transmission Protocol", Internet Draft draft-coene-sctp-multihome-03.txt, February 2002.

[16] Bagnulo, M., Garcia-Martinez, A. and I. Soto, "Application of the MIPv6 protocol to the multi-homing problem", Internet Draft draft-bagnulo-multi6-mnm-00, February 2003.

[17] Tattam, P., "Preserving active TCP sessions on Multihomed IPv6 Networks", Internet Draft , URL http://jazz-1.trumpet.com.au/ipv6-draft/preserve-tcp.txt, August 2001.

[18] Bradner, S., "Key words for use in RFCs to indicate requirement levels", RFC 2119, March 1997.

[19] <http://bgp.potaroo.net>

Authors' Addresses

Cedric de Launois
UCL/INGI
Place Ste-Barbe 2
B-1348 Louvain-la-Neuve
Belgium

EMail: delaunois@info.ucl.ac.be
URI: http://www.info.ucl.ac.be/people/delaunoi/

Olivier Bonaventure
UCL/INGI
Place Ste-Barbe 2
1348 Louvain-la-Neuve
Belgium

EMail: bonaventure@info.ucl.ac.be
URI: http://www.info.ucl.ac.be/people/OBO/

Intellectual Property Statement

Full Copyright Statement

Acknowledgement

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