INTERNET-DRAFT

K. Ohira
M. Kozuka
Y. Okabe
Kyoto University
Y. Koyama
Trans New Technology
July 19, 2004

IPv6 Address Assignment and Route Selection for End-to-End Multihoming

<draft-ohira-assign-select-e2e-multihome-03.txt>

Status of this Memo

Copyright Notice

draft-ohira-assign-select-e2e-multihome-03.txt July 19, 2004

Abstract

In this document, we propose a way of address assignment and route selection suitable for "Host-Centric" multihoming, where an end node plays the main role of multihoming.

The key techniques are a hierarchical address assignment and a source address dependent routing (SADR) for only default route entry.

In our proposal, an IP address itself has some sense of routing information. We propose that an address assignment SHOULD be hierarchical. Under the conditions that address assignment is hierarchical, when someone delegates an address block, it means that it also hands routing information to its downstream at the same time. In this manner, a host which has several addresses can select which upstream to go through with by selecting its source address.

1. Introduction

As described in RFC3582 [3582], the main purpose of multihoming is getting redundancy and load sharing.

Usually, if someone wants a site to be multihomed, he has to get an AS number and connect to upstream ISP with BGP peering. This method is too difficult to configure and AS numbers are limited, so it is actually unable for small sites to be multihomed.

In the BGP method, we MUST trust any routing information from outside of a site and reconfigure routing information inside of the site, so this method is problematic at reliability and its speed.

Furthermore, in IPv6, it is REQUIRED that we aggregate addresses more thoroughly than in IPv4 because of its address length.

In this document, we show a scalable way to multihome with a very little information from outside. It is based on so-called "Host-Centric" [HOSTS] multihome, where an end node plays the main role.

2. Types of multihoming sites

2.1 Definition of Terms

Terms which are not referred here are used as the meaning in RFC3582 [3582].

The term "site" means a small site such as a home network.

The term "end-to-end" means that an identity / locator binding should be done within the end-to-end transport protocol layer (Section 5.3.3 of [ARCHITECTURE]) though this proposal is applicable with an IP layer solution or a shim layer solution.

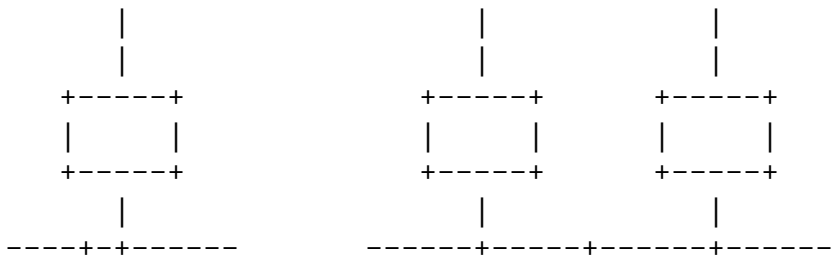
2.2 Network topology of a site

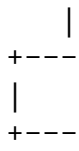
First, it should be noted that every host centric solutions in a small site is completed within the site. In other words, it is nothing to do with an action of an upstream ISP.
Moreover, even if another site adopts another solution, a communication between those two sites must not inhibited only because of it.

Furthermore, in the most of usual methods, routing information in a site deeply depends on that from outside of the site. This is problematic at the point of reliability and its speed.

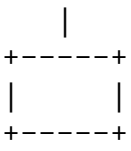
A site is classified into the following 3 cases according to the location(s) of site exit router(s):

1. Single link, single exit router (Fig. 1),
2. Single link, multiple exit routers (Fig. 2) and
3. Multiple link, multiple exit routers (Fig. 3).

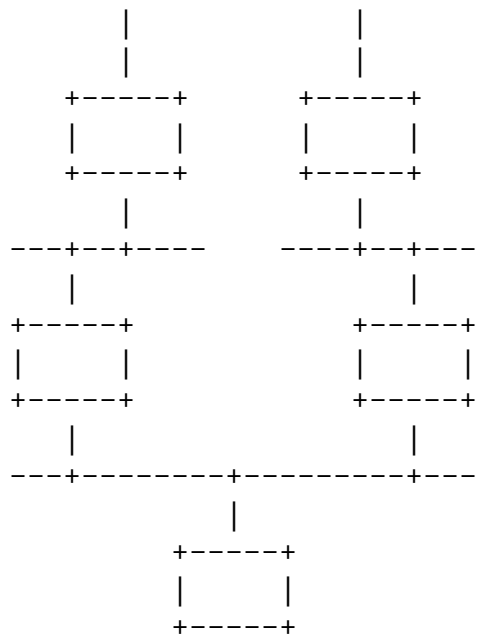




[Fig. 1]



[Fig. 2]



[Fig. 3]

At this time, we assume that a site is assigned PA (Provider Aggregatable) address prefixes from each upstream (multi-addressing) and each upstream carries out ingress filtering on an advice of RFC2267 [2267].

In the case of 1, if the exit router has a mechanism of forwarding a packet to the proper upstream, the other nodes in the site do not have to be modified at all.

In the case of 2, if the exit routers have a mechanism of forwarding a packet which should go through with another exit router to the proper exit router, the other nodes in the site do not have to be modified at all.

In this draft, we will mainly discuss about the case 3.

3. Analysis of source address dependent routing solutions

[HOSTS] proposes a solution that setting tunnels between site exit routers and that a site exit router forward a packet to the proper one according to its source address. With this method, each host (or each end on a host) in a multihomed site can select an upstream ISP respectively.

This tunnel solution is classified as a kind of source address dependent routing (refer to the Section 4.2 of [HOSTS].) This

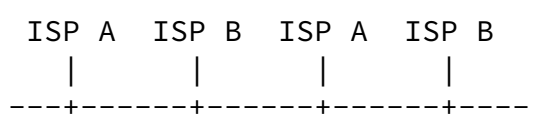
solution need modification only on site exit routers but there is a fear that a path from a source host to a proper site exit router may not be shortest.

Therefore, we pursue of a method which is good at fault tolerance and which can route an outgoing packet to a proper site exit router with shortest path even in a multi-link and multi-exit-router site.

Here, we classify a multi-link and multi-exit-router site into the following two cases:

a) A site exit router to an ISP (global prefix) and
b) Multiple site exit routers to an ISP (global prefix.)

Compared the case a with the case b, it is considered that the case b is more generalized. Therefore, we consider the case b. An example is as shown in Fig. 4.



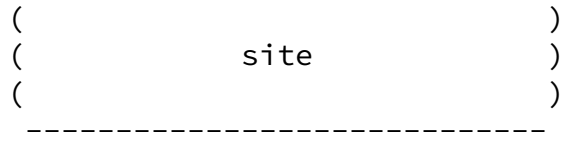


Fig. 4: Multiple site exit routers to an ISP

At a site as shown in Fig. 4, not only a selection which upstream ISP a packet should be goes through with but also a selection which exit router it should be go through with is needed.

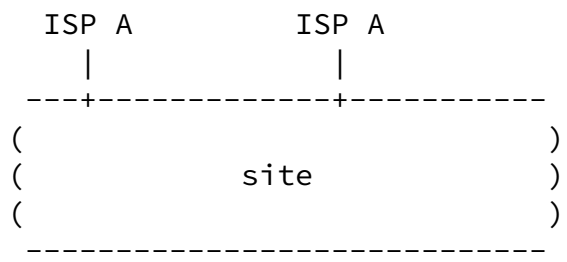


Fig. 5: Multiple site exit routers to the same ISP

In this section, in order to gurarantee a path from a source host to a site exit router the shortest, we discuss about adding new

features to either routers or hosts in a site.

3.1 Host enhancement solution

As a solution that a source host selects a proper site exit router, following solutions are considered:

i) IP in IP tunneling between a source host and a site exit router and
ii) Using a routing option header.

These solutions do not need any enhancement on routers in a site except site exit routers. Moreover, by using these solutions with tunneling between site exit routers, all hosts do not need these

solutions. However, these solutions have a drawback that they need to expand the size of IP packet.

These solutions request a source host to select not only an upstream ISP but also a site exit router to the ISP which go through with.

3.2 Router enhancement solution

A multihomed site can be divided into 2 fields as shown in Fig. 6.

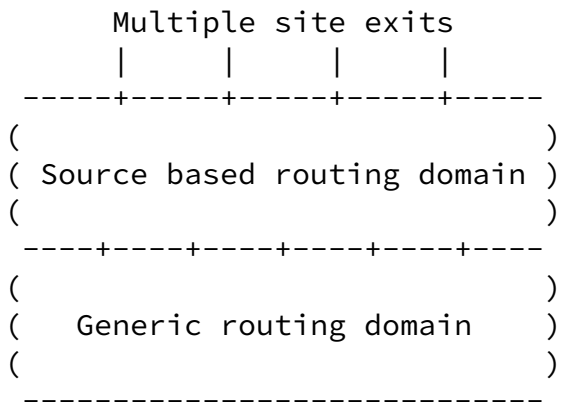


Fig. 6: Source based routing domain
(Cited from Section 4.2 of [HOSTS])

Here, a solution that every router in the source based routing domain holds source address dependent routing policy is considered.

This solution does not need any enhancement on a host in a site. Moreover, this solution does not expand the size of an IP packet.

However, for effectiveness, it is preferable that the most part of a site is source based routing domain. In other words, the most routers of the site are required to be enhanced. This may become a drawback of this solution.

With this solution, a source host can select an upstream ISP and/or a site exit router to the ISP which go through with.

4. Proposed solution

In order to solve issues described at the previous section, we propose the following method.

4.1 Network topology and Hierarchical Address assignment

This method is categorized into multi-addresses model. In this model, a site is delegated a prefix of PA address from each upstream.

At this time, in order to be easy for a site to construct dynamic network:

- The length of the prefix which an ISP delegates to a site is equal to or less than 48.

A site exit router advertises to the all nodes in the site:

- Delegated prefix(es) and
- Information that this exit router is proper one for packets with such prefixed address as source address.

In order to advertise, RR (Router Renumbering) [2894], DHCPv6 [3315] Prefix Option [3633] and/or RA (Router Advertisement) [2461] may be useful.

An usual configuration is as below.

- Upper n bits: Global prefix or some temporary "site-local" prefix such as Unique Local IPv6 Unicast Address [LOCALADDR].
- Middle m bits: With manual configuration or auto configuration such as zeroconf [2462, STATELESS], REQUIRED to complete configurations independently of the upper n bit.
- Lower 128-n-m bits: Node identifier.

Where n and m are in conformity with the definition in RFC3513

[3513].

As for an issue of association between an address assignment and the SADR in a site, see [MULTILINK].

4.2 Route Selection

With an usual method, routing information inside of a multihomed site depends on that of outside, then there are some problems:

- A node in the site has to trust the unreliable information.
- When a connection between site exit router and upstream fails, it spends a few minutes to recover (some connections may timeout).

In our proposing method,

- In a site, route information of only inside of the site is advertised.
- Route information about outside of the site is only default route in order to rely on the least information about outside.

In order to select the proper exit router, nodes SHOULD refer the source address of a packet bound for the outside of the site. In other words,

- Source Address Dependent Routing (SADR) SHOULD be done for default route entries.

In our method, an external route is expressed as a connection to the "next-hop" upstream. Therefore, this information is reliable as information about inside of the site. Besides, because the whole (or a part of) information about connections to upstream is advertised to all nodes in the site and a node (or an application running on the node) can select or change its source address, a more rapid change routes is expected.

4.3 Site Exit Selection

In the section 3, we discussed about solutions that a source host selects a site exit router depending on a source address and that a path from the source host to the exit router is the shortest one.

In the case of Fig. 5, it is difficult to make some rule whether exit router to go through with only from source address. Therefore, a solution which does not require a host to select a site exit router than a solution which requires.

In other words, a selection which upstream ISP to go through with should be done by a source host but a selection which exit router to go through with to the ISP should be done by routing mechanism in the site.

As for this, this draft proposes a solution that all routers in a site carry out source address dependent routing (SADR.)

With the proposed solution, even a legacy host can send a packet to a proper site exit router with the shortest path.

The proposed solution requires site exit routers and routers around site exit (details is described in 4.3.1) to enhance. Because of it, it may seem difficult to apply a site as a short term solution. However, because this solution is able to be adopted gradually, this solution can be a candidate of long term solutions.

Conventionally, a source address dependent routing is considered to have some drawbacks as below:

1) A packet forwarding may be looped,
2) The size of routing table at a router may become too large and
3) Massive re-engineering of routers and routing protocol may be needed.

We show that these anxiety are disappeared by limiting the target routes of source address dependent routing in 4.3.1, 4.3.2 and 4.3.3 respectively.

Therefore, a source address dependent routing becomes a solution which can be used at small site.

4.3.1 Avoidance of packet forwarding loop

At a site as shown in Fig. 6, source based routing domain must not be discrete. This is because the following reason.

In a site with discrete source based routing domains as shown in Fig. 7, when a host in generic routing domain send a packet and the packet reaches at the left hand side of source based routing domain, if the proper site exit router is in the right hand side of source based routing domain, then the packet is forwarded around the border between the left hand side of source based routing domain and generic routing domain until TTL becomes 0.

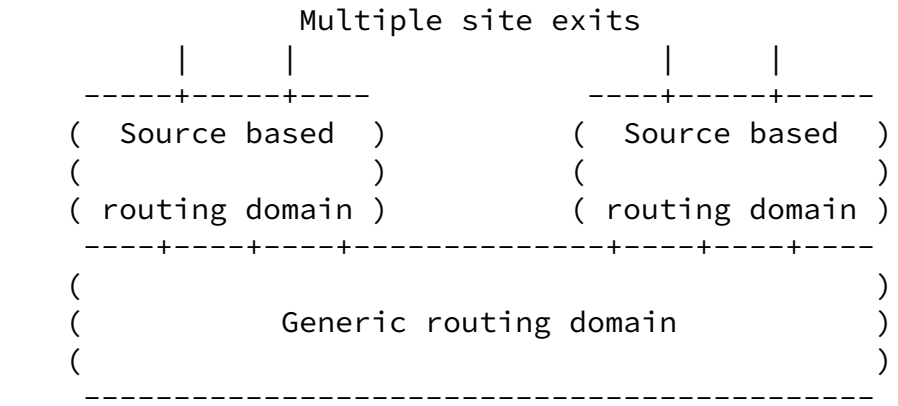


Fig. 7: A discrete source based routing domains

However, even if there are some physically discrete source based routing domains in a site, if these domains are logically connected with tunnels, the problem described above is not occurred.

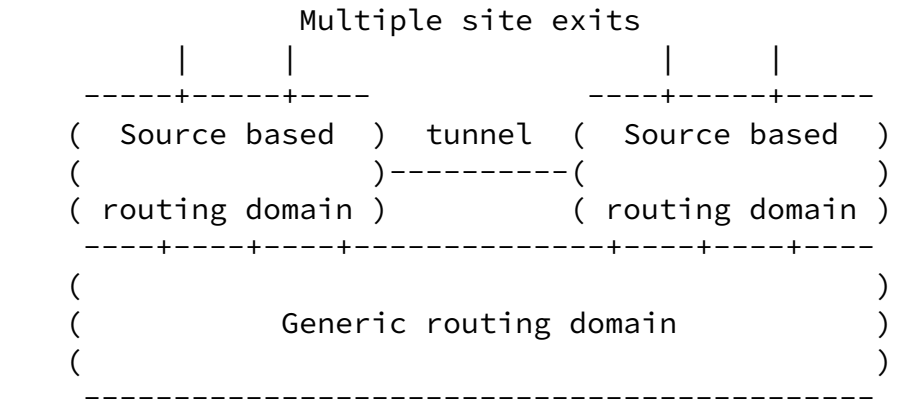


Fig. 8: Concatenated source based routing domains with a tunnel

Here, it should be noted that this tunnel itself goes through with the generic routing domain, so that an capsuling packet itself is routed according with the destination address of it.

4.3.2 Suppression of the amount of routing information

The purpose of this solution is forwarding a packet for a host which is outside of a site from a source host to a proper site exit router according to the source address of the packet through the shortest path.

Based on the discussion in 4.3.1, even in the source based routing

domain, a packet which destination is inside the site must be routed source address independently.

A small site seems hardly to get detailed routing information about outside of the site and the site exit router mostly gets only the information about the address of counter router in an upstream ISP as the default gateway of the site. Moreover, the site is not considered to transit any traffic between two third parties. Therefore, a limitation route outside of the site to default route does not seem to occur any special problems.

This draft proposes to limit route outside of the site to default route (dst=::/0) and to apply source address dependent routing only for default route entry. These are useful to suppress the amount of routing information.

If a site exit router gets detailed routing information about outside of the site, it can be used as base information of source address selection by a source host. However, this issue is outside of the scope of this draft.

4.3.3 Suppression of impact on a router or a routing protocol

This proposed solution can be carried out without any change of packet format of routing information advertised within a site. Especially in a link state routing protocol such as OSPF, because routing information is not merged normally, information about multiple default routes can be announced in a site respectively.

In a distance vector routing protocol such as RIP, routing information about the same destination is merged. Therefore, when a

site exit router announces a default route within the site, the destination of each information should not be as the same kind (like ::/0) but should be different for each global prefix assigned by an upstream ISP. An example is described in Appendix B. However, as in a link state routing protocol, the packet format of routing information does not need any changes.
In this case, it is considered that each default route entry is expressed as 'the whole address space which a site is assigned from an ISP.'

As explained above, the format of routing information does not need any change. All what is required is that a router which received information about default route should hold default route information for each global prefix respectively.

In order to carry out it, enhancement to every router to keep default

route information for each global prefix respectively or not to merge information about default route is required.

An implementation of this solution is described in Appendix A.

5. Evaluation of this proposal and multihoming goals

In this section, we will evaluate how much our proposing method meets the requirements pointed out in RFC3582 [3582].

5.1 Capabilities of IPv4 Multihoming

5.1.1 Redundancy

Every external connection is treated completely separate. Therefore, our proposing method is able to continue a connection unless all external connection fail.

5.1.2 Load Sharing

Each end of transport layer is able to distribute both inbound and outbound traffic between multiple transit providers.

(cf. In host centric multihoming, each host is able to distribute.)

5.1.3 Performance

No information between upstream ISPs is REQUIRED.

If a corresponding node can divide a stream into several destination addresses, we can accomplish to distribute inbound traffic.

5.1.4 Policy

Policies of a site will be expressed as ingress/egress filtering rules. If a site does not want a host to use an external connection, the site can neglect to re-delegate an address with the prefix specific to the external connection.

5.1.5 Simplicity

Our proposing method is very simple. In the simplest case, we may be able to configure with only one command or jumper switch.

5.1.6 Transport-Layer Survivability

With the method described in the section 3, we can select a route to use from several candidates.

At a point of view of an application, we expect that all messages are exchanged in just one connection. In order to meet this requirement, we need some co-operation with L4 (for UDP, it may be L7).

There are two approaches as below:

- Define an intermediate layer between the transport layer and IP layer. In this approach, the intermediate layer merges several IP addresses into one transport layer address, so a transport layer protocol thinks that an unchanged connection continues.

- Transport layer protocol itself directly handles multiple IP addresses.

As an intermediate layer, LIN6 [LIN6], HIP [HIP, HIP-MM], MAST [MAST],
SIM [SIM] and etc. are proposed.

In order to bind several IP address to a TCP connection, the TCP extension to multihome [TCP-MH] is proposed. SCTP [SCTP, SCTP-MH] can handle several IP addresses in an association.

A connection between a node and another node may have several paths (i.e. several pairs of source/destination addresses). Decision of priority of these pairs SHOULD be done in L4/L7 with following the rules described in RFC3484 [3484].

However, these methods are both disputable about how to hold binding relation and its security issues.

5.1.7 Impact on DNS

Any change of external connections of a site cause change(s) of prefixes which the site has. Therefore, in the worst case, we may be required to change DNS information at every time.

5.1.8 Packet Filtering

Our proposing method is designed to co-operate with ingress/egress filtering. If the source address of an IP packet is valid then the packet is forwarded to the proper next hop, otherwise the packet will

be discarded.

5.2 Additional Requirements

5.2.1 Scalability

Only a Provider Aggregatable IP address block from upstream is REQUIRED. This address is always aggregated at upstream, so even if the number of multihoming site with our proposing method increase, the number of routing information at DFZ (Default Free Zone). Still

more, no AS number is REQUIRED for a site to be multihomed.

In these points, our proposing method is very scalable.

5.2.2 Impact on Routers

The SADR is REQUIRED for at least one router in a multihoming site. If there are some routers which cannot handle SADR, according to the position, routing loop may be occurred.

The authors think that this requirement is relatively little because the SADR is required only for default route entry.

These modifications do not prevent normal single-homed operations. In a single-homed site, modified routers and unmodified routers can coexist.

5.2.3 Impact on Hosts

The SABR is REQUIRED for all end hosts who want to be fully multihomed. However, a legacy (without SADR) host can be obtain some functions of multihome.

If you want to bind several IP addresses to a single TCP connection, TCP Extension for Multihoming may be useful.

5.2.4 Interaction between Hosts and the Routing System

No interactions are REQUIRED except for information about proper next hop for each source address prefixes.

5.2.5 Operations and Management

Administrators of a site are completely capable to monitor the state or to configure parameters of multihoming. At this time, the administrators do not have to do any co-operative work with administrators of upstream.

5.2.6 Cooperation between Transit Providers

Our proposing method does not require any co-operative work between upstream providers at all.

5.2.7 Multiple Solutions

In a single network segment, our proposing method is RECOMMENDED to be used solely. However, we can divide the site into two or more segment in order to use those multiple solutions respectively.

6. Things multi6 developers should think about

In this section, we will evaluate our proposing method at the point of questions described in [THINKABOUT].

6.1 On the wire behavior

6.1.1 How will your solution solve the multihoming problem?

Put source address dependent routing policy on every router in a site.

6.1.2 At what layer is your solution applied, and how?

At the layer 3. The proposal is applied in every packet. The source address field is used.

6.1.3 Why is the layer you chose the correct one?

Because this proposal only relates to the issue of packet forwarding from a source host to a proper site exit router according to the source address of the packet.

6.1.4 Does your solution address mobility?

No. This proposal is orthogonal to MOBILEIP-V6.

6.1.5 Does your solution expand the size of an IP packet?

No.

6.1.6 Will your solution add additional latency?

No.

6.1.7 Can multihoming capabilities be negotiated end to end during a connection?

No.

6.1.8 Do you change the way fragmenting is handled?

N/A.

6.1.9 Are there any layer 2 implications to your proposal?

No.

6.2 Identifiers and locators

N/A.

6.3 Routing system interactions

6.3.1 Does your solution change existing aggregation methods?

No.

6.3.2 If you introduce any new name spaces, do they require aggregation?

No.

6.3.3 Are there any changes to ICMP error semantics?

No.

6.4 Name service interactions

N/A.

6.5 Application concerns and backward compatibility

6.5.1 What application/API changes are needed?

No change is needed.

6.5.2 Is this solution backward compatible with "old" IP version 6?

An user of a normal IPv6 node may receive an ICMP redirect message.

6.5.3 Is your solution backward compatible with IPv4?

This proposal is orthogonal to 6to4 gateways. This mechanism does not consider IPv4.

6.5.4 Can IPv4 devices take advantage of this solution?

Yes.

6.5.5 What is the impact of your solution on different types of sites?

Single homed sites:	No impacts.
Small multihomed sites:	The main target of this proposal. Source address dependent routing policy on each router in a site.
Large multihomed sites:	May prefer another solution.

Ad-hoc sites: Not preferable.

Short lived connections: No impacts.

6.5.6 How will your solution interact with other middleboxes?

We do not use any middlebox.

6.5.7 Referrals

This proposal does not make it worse.

6.6 Legal concerns

N/A.

7. Security Considerations

This proposal does not introduce any new kind of messages. Therefore, the authors think this proposal make it less secure than the current situation.

8. IANA Considerations

There are no IANA considerations in this document.

9. Acknowledgements

The authors thank Mr. Arifumi Matsumoto who grants us permission to publish an implementation of SADR for default route entry.

The downloads page of the implementation is noted at Appendix A.

Appendix A: Implementations of SADR for default route entry

SADR for default route entry can be put in practice with some extensions of policy routing.

We can get it with ipfilter (ipf,) and we have another implementation of extended ECMP which runs on NetBSD (KAME.)

The implementation of extended ECMP is now available at http://www.rd.miako.net/~arifumi/ietf/kame-20031020-netbsd161-ecmpsabr.diff.

Appendix B: An expression of a default route at a distance vector routing protocol

For example, a site is assigned an address space A::/48 from the ISP A and an address space B::/48 from the ISP B.

In this case, a site exit router a which is connected to the ISP A announces routing information of dst=A::/48 within the site with distance vector routing protocol such as RIP.

If all routers in the site know that the site is assigned A::/48 beforehand, at a router which receive the above information can recognize that the information must be treated as the default route in the site. The same is true in B::/48.

By using this replacement expression, even with DV routing, multiple default routes can be announced in demultiplexable style.

Here, the reason why A::/48 and B::/48 are used as alternative expression of ::/0 is that they are never used in the site because of longest match rule of routing information.

References

[3582] J. Abley, et. al., "Goals for IPv6 Site-Multihoming Architectures", RFC3582, August 2003.

[SCTP-MH] L. Coene, et. al., "Multihoming: the SCTP solution",

Internet Draft (work in progress), draft-coene-multi6-sctp-00.txt, January 2004.

[2894] M. Crawford, "Router Renumbering for IPv6", RFC2894, August 2000.

[MAST] D. Crocker, "Multiple Address Service for Transport (MAST): An Extended Proposal", Internet Draft (work in progress), draft-crocker-mast-proposal-01.txt, September 2003.

[3484] R. Draves, "Default Address Selection for Internet Protocol version 6 (IPv6)", RFC3484, February 2003.

[3315] R. Droms, Ed., et. al., "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC3315, July 2003.

[2267] P. Ferguson, et. al., "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", RFC2267, January 1998.

[LOCALADDR] R. Hinden, et. al., "Unique Local IPv6 Unicast Address", Internet Draft (work in progress), draft-hinden-ipv6-

global-local-addr-02.txt, June 2003.

[HOSTS] C. Huitema, et. al., "Host-Centric IPv6 Multihoming", Internet Draft (work in progress), draft-huitema-multi6-hosts-03.txt, February 2004.

[ARCHITECTURE] G. Huston, "Architectural Approaches to Multi-Homing for IPv6", Internet Draft (work in progress), draft-ietf-multi6-architecture-00.txt, July 2004.

[THINKABOUT] E. Lear, "Things MULTI6 Developers should think about", Internet Draft (work in progress), draft-ietf-multi6-things-to-think-about-00.txt, June 2004.

[TCP-MH] A. Matsumoto, et. al., "TCP Multi-Home Options", Internet Draft (work in progress), draft-arifumi-tcp-mh-00.txt, October 2003.

[HIP] R. Moskowitz, et. al., "Host Identity Protocol", Internet Draft

(work in progress), draft-moskowitz-hip-09.txt, February 2004.

[2461] T. Narten, et. al., "Neighbor Discovery for IP Version 6 (IPv6)", RFC2461, December 1998.

[HIP-MM] P. Nikander, et. al., "End-Host Mobility and Multi-Homing with Host Identity Protocol", Internet Draft (work in progress), draft-nikander-hip-mm-02.txt, July 2004.

[SIM] E. Nordmark, "Strong Identity Multihoming using 128 bit Identifiers (SIM/CBID128)", Internet Draft (work in progress), draft-nordmark-multi6-sim-01.txt, October 2003.

[MULTILINK] K. Ohira, et. al., "Hierarchical IPv6 Subnet ID Autoconfiguration for Multi-Address Model Multi-Link Multihoming Site", Internet Draft (work in progress), draft-ohira-multi6-multilink-auto-prefix-assign-00.txt, January 2004.

[SCTP] R. Stewart, et. al., "Stream Control Transmission Protocol", RFC2960, October 2000.

[LIN6] F. Teraoka, et. al., "LIN6: A Solution to Multihoming and Mobility in IPv6", Internet Draft (work in progress), draft-teraoka-multi6-lin6-00.txt, December 2003.

[2462] S. Thomson, et. al., "IPv6 Stateless Address Autoconfiguration", RFC2462, December 1998.

[3633] O. Troan, et. al., "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC3633, December 2003.

[STATELESS] T. Yoshihiro, et. al., "Stateless Autoconfiguration of IPv6 Site-Local Address." Communications and Computer Networks pp.. 299--304, IASTED (2002)

Authors' Addresses

Kenji Ohira

Graduate School of Informatics
Kyoto University
Yoshida-Hommachi, Sakyo-ku, Kyoto 606-8501 JAPAN
Tel: +81-75-753-7468
Fax: +81-75-753-7472
Email: ohira@net.ist.i.kyoto-u.ac.jp

Masahiro Kozuka
Faculty of Law, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501 JAPAN
Tel: +81 75-753-7468
Fax: +81 75-753-7472
Email: ma-kun@kozuka.jp

Yasuo Okabe
Academic Center for Computing and Media Studies
Kyoto University
Yoshida-Hommachi, Sakyo-ku, Kyoto 606-8501 JAPAN
Tel: +81-75-753-7458
Fax: +81-75-751-0482
Email: okabe@i.kyoto-u.ac.jp

Youichi Koyama
Trans New Technology, Inc.
Inohara BLDG. 2F, 72 Tanaka Monzencho, Sakyo,
Kyoto 606-8225 JAPAN
Tel: +81-75-706-9800
Fax: +81-75-706-9801
Email: koyama-y@trans-nt.com

Intellectual Property Statement

Copies of IPR disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr.

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

Copyright Statement

Acknowledgment

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