Network Working Group A. Matsumoto
Internet-Draft J. Kato
Intended status: Standards Track T. Fujisaki
Expires: December 12, 2011 NTT
June 10, 2011

Update to RFC 3484 Default Address Selection for IPv6
draft-ietf-6man-rfc3484-revise-03.txt

Abstract

RFC 3484 describes algorithms for source address selection and for destination address selection. The algorithms specify default behavior for all Internet Protocol version 6 (IPv6) implementations. This document specifies a set of updates that modify the algorithms and fix the known defects.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction 3
2. Specification 3
  2.1. Changes related to the default policy table 3
    2.1.1. ULA in the policy table 4
    2.1.2. Teredo in the policy table 4
    2.1.3. 6to4, Teredo, and IPv4 prioritization 4
    2.1.4. Deprecated addresses in the policy table 5
    2.1.5. Renewed default policy table 5
  2.2. The longest matching rule 5
  2.3. Utilize next-hop for source address selection 6
  2.4. Private IPv4 address scope 6
  2.5. Deprecation of site-local unicast address 7
3. Security Considerations 7
4. IANA Considerations 7
5. References 7
  5.1. Normative References 7
  5.2. Informative References 8
Appendix A. Acknowledgements 9
Appendix B. Past Discussion 9
  B.1. The longest match rule 9
  B.2. NAT64 prefix issue 10
Appendix C. Revision History 10
Authors' Addresses 11

1. Introduction

The IPv6 addressing architecture [RFC4291] allows multiple unicast addresses to be assigned to interfaces. Because of this IPv6 implementations need to handle multiple possible source and destination addresses when initiating communication RFC 3484 [RFC3484]. specifies the default algorithms, common across all implementations, for selecting source and destination addresses so that it is easier to predict the address selection behavior.

After RFC 3484 was specified, some issues have been identified with the algorithm specified there. The issues are related to the longest match algorithm used in Rule 9 of Destination address selection breaking DNS round-robin techniques, and prioritization of poor IPv6 connectivity using transition mechanisms over native IPv4 connectivity.

There have also been some significant changes to the IPv6 addressing architecture that require changes in the RFC 3484 policy table. Such changes include the deprecation of site-local unicast addresses [RFC3879] and the IPv4-compatible IPv6 addresses, the introduction of Unique Local Addresses [RFC4193] etc.

This document specifies a set of updates that modify the algorithms and fix the known defects.

2. Specification

2.1. Changes related to the default policy table

The default policy table is defined in RFC 3484 Section 2.1 as follows:

```
      Prefix        Precedence Label
      ::1/128               50     0
```

::/0	40	1
2002::/16	30	2
::/96	20	3
::ffff:0:0/96	10	4

The changes that should be included into the default policy table are those rules that are universally useful and do no harm in every reasonable network envionment. The changes we should consider for the default policy table are listed in this sub-section.

The policy table is defined to be configurable. The changes that are useful not universally but locally can be put into the policy table

manually or by using the auto-configuration mechanism proposed as a DHCP option [I-D.ietf-6man-addr-select-opt].

2.1.1. ULA in the policy table

RFC 5220 [RFC5220] Section 2.1.4, 2.2.2, and 2.2.3 describes address selection problems related to ULA [RFC4193]. These problems can be solved by either changing the scope of ULA to site-local, or by adding an entry for default policy table entry that has its own label for ULA.

Centrally assigned ULA [I-D.ietf-ipv6-ula-central] is proposed, and is assigned fc00::/8. Using the different labels for fc00::/8 and fd00::/8 makes sense if we assume the same kind of address block is assigned in the same or adjacent network. However, we cannot expect that the type of ULA address block and network adjancency commonly have any relationships.

Regarding the scope of ULA, ULA has been specified with a global scope because the reachability of the ULA was intended to be restricted by the routing system. Since the ULAs will not be exposed outside of its reachability domain, if an ULA is available as a candidate destination address, it can be expected to be reachable.

if we change the scope of ULA smaller than global, we can prioritize ULA to ULA communication over GUA to GUA communication. At the same time, however, finer-grained configuration of ULA address selection will be impossible. For example, even if you want to priorize communication related to the only /48 ULA prefix used in your site, and do not want to prioritize communication to any other ULA prefix, such a policy cannot be implemented in the policy table. So, this draft proposes the use of the policy table to differentiate ULA from GUA.

2.1.2. Teredo in the policy table

Teredo [RFC4380] is defined and has been assigned 2001::/32. This address block should be assigned its own label in the policy table. Teredo's priority should be less or equal to 6to4, considering its characteristic of transitional tunnel mechanism. About Windows, this is already in the implementation.

2.1.3. 6to4, Teredo, and IPv4 prioritization

Regarding the prioritization between IPv4 and these transitional mechanisms, the connectivity of them are known to be worse than IPv4. These mechiansms are said to be the last resort access to IPv6 resources. While 6to4 should have higher precedence over Teredo, in

that 6to4 host to 6to4 host communication can be over IPv4, which can result in more optimal path, and 6to4 does not need NAT traversal.

2.1.4. Deprecated addresses in the policy table

IPv4-compatible IPv6 address (::/96) is deprecated [RFC4291]. IPv6 site-local unicast address (fec0::/10) is deprecated [RFC3879]. 6bone testing address [RFC3701] was returned.

These addresses were removed from the current specification. So, it should not be treated differently, especially if we plan future re-use of these address blocks. Hense, 6bone testing address block should not be treated specially.

Considering the inappropriate use of these address blocks especially in outdated implementations and bad effects brought by them, it should be labeled differently from the legitimate address blocks as far as the address block is reserved by IANA.

2.1.5. Renewed default policy table

After applying these updates, the default policy table will be:

Prefix	Precedence	Label
::1/128	60	0
fc00::/7	50	1
::/0	40	2
::ffff:0:0/96	30	3
2002::/16	20	4
2001::/32	10	5
::/96	1	10
fec0::/10	1	11

2.2. The longest matching rule

This issue is related to the longest matching rule, which was found by Dave Thaler. It is malfunction of DNS round robin technique. It is common for both IPv4 and IPv6.

When a destination address DA, DB, and the source address of DA Source(DA) are on the same subnet and Source(DA) == Source(DB), DNS round robin load-balancing cannot function. By considering prefix lengths that are longer than the subnet prefix, this rule establishes preference between addresses that have no substantive differences between them. The rule functions as an arbitrary tie-breaker between the hosts in a round robin, causing a given host to always prefer a given member of the round robin.

By limiting the calculation of common prefixes to a maximum length equal to the length of the subnet prefix of the source address, rule 9 can continue to favor hosts that are nearby in the network hierarchy without arbitrarily sorting addresses within a given network. This modification could be written as follows:

Rule 9: Use longest matching prefix.

When DA and DB belong to the same address family (both are IPv6 or both are IPv4): If CommonPrefixLen(DA & Netmask(Source(DA)), Source(DA)) > CommonPrefixLen(DB & Netmask(Source(DB)), Source(DB)), then prefer DA. Similarly, if CommonPrefixLen(DA & Netmask(Source(DA)), Source(DA)) < CommonPrefixLen(DB & Netmask(Source(DB)), Source(DB)), then prefer DB.

2.3. Utilize next-hop for source address selection

The RFC 3484 source address selection rule 5 defines the address that is attached to the outgoing interface should be preferred as the source address. This rule is reasonable considering the prevalence of Ingress Filtering described in BCP 38 [RFC2827]. This is because an upstream network provider usually assumes it receives those packets from their customer that have the delegated addresses as the source addresses.

This rule, however, is not effective in such a environment described in RFC 5220 Section 2.1.1, where a host has multiple upstream routers on the same link and has addresses delegated from each upstream routers on single interface.

So, a new rule 5.1 that utilizes next-hop information for source address selection is inserted just after the rule 5.

Rule 5.1: Use an address assigned by the selected next-hop.

If SA is assigned by the selected next-hop that will be used to send to D and SB is assigned by a different next-hop, then prefer SA. Similarly, if SB is assigned by the next-hop that will be used to send to D and SA is assigned by a different next-hop, then prefer SB.

2.4. Private IPv4 address scope

When a packet goes through a NAT, its source or destination address can get replaced with another address with a different scope. It follows that the result of the source address selection algorithm may be different when the original address is replaced with the NATed address.

The algorithm currently specified in RFC 3484 is based on the assumption that a source address with a small scope cannot reach a destination address with a larger scope. This assumption does not hold if private IPv4 addresses and a NAT are used to reach public IPv4 addresses.

Due to this assumption, in the presence of both NATed private IPv4 address and transitional addresses (like 6to4 and Teredo), the host will choose the transitional IPv6 address to access dual-stack peers [I-D.denis-v6ops-nat-addrsel]. Choosing transitional IPv6 connectivity over native IPv4 connectivity is not considered to be a very wise result.

This issue can be fixed by changing the address scope of private IPv4 addresses to global. Such change has already been implemented in some OSes.

2.5. Deprecation of site-local unicast address

RFC 3484 contains a few "site-local unicast" and "fec0::" description. It's better to remove examples related to site-local unicast address, or change examples to use ULA. Possible points to be re-written are below.

- 2nd paragraph in RFC 3484 Section 3.1 describes scope comparison mechanism.
- RFC 3484 Section 10 contains examples for site-local address.

3. Security Considerations

No security risk is found that degrades RFC 3484.

4. IANA Considerations

Address type number for the policy table may have to be assigned by IANA.

5. References

5.1. Normative References

[RFC1794] Brisco, T., "DNS Support for Load Balancing", RFC 1794, April 1995.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and

E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

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

[RFC3701] Fink, R. and R. Hinden, "6bone (IPv6 Testing Address Allocation) Phaseout", RFC 3701, March 2004.

[RFC3879] Huitema, C. and B. Carpenter, "Deprecating Site Local Addresses", RFC 3879, September 2004.

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, February 2006.

[RFC4380] Huitema, C., "Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs)", RFC 4380, February 2006.

[RFC5220] Matsumoto, A., Fujisaki, T., Hiromi, R., and K. Kanayama, "Problem Statement for Default Address Selection in Multi-Prefix Environments: Operational Issues of RFC 3484 Default Rules", RFC 5220, July 2008.

5.2. Informative References

[I-D.chown-addr-select-considerations]
Chown, T., "Considerations for IPv6 Address Selection Policy Changes", draft-chown-addr-select-considerations-03 (work in progress), July 2009.

[I-D.denis-v6ops-nat-addrsel]
Denis-Courmont, R., "Problems with IPv6 source address selection and IPv4 NATs", draft-denis-v6ops-nat-addrsel-00 (work in progress), February 2009.

[I-D.ietf-6man-addr-select-opt]
Matsumoto, A., Fujisaki, T., and J. Kato, "Distributing Address Selection Policy using DHCPv6", draft-ietf-6man-addr-select-opt-00 (work in progress), December 2010.

[I-D.ietf-ipv6-ula-central]
Hinden, R., "Centrally Assigned Unique Local IPv6 Unicast Addresses", draft-ietf-ipv6-ula-central-02 (work in

progress), June 2007.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.

[RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.

Appendix A. Acknowledgements

Authors would like to thank to Dave Thaler, Pekka Savola, Remi Denis-Courmont and the members of 6man's address selection design team for their invaluable contributions to this document.

Appendix B. Past Discussion

This section summarizes discussions we had before related to address selection mechanisms.

B.1. The longest match rule

RFC 3484 defines that the destination address selection rule 9 should be applied to both IPv4 and IPv6, which spoils the DNS based load balancing technique that is widely used in the IPv4 Internet today.

When two or more destination addresses are acquired from one FQDN, the rule 9 defines that the longest matching destination and source address pair should be chosen. As in RFC 1794, the DNS based load balancing technique is achived by not re-ordering the destination addresses returned from the DNS server. The Rule 9 defines deterministic rule for re-ordering at hosts, hence the technique of RFC 1794 is not available anymore.

Regarding this problem, there was discussion in IETF and other places like below.

Discussion: The possible changes to RFC 3484 are as follows:

1. To delete Rule 9 completely.
2. To apply Rule 9 only for IPv6 and not for IPv4. In IPv6, hiearchical address assignment is general principle, hence the longest matchin rule is beneficial in many cases. In IPv4, as

stated above, the DNS based load balancing technique is widely used.

3. To apply Rule 9 for IPv6 conditionally and not for IPv4. When the length of matching bits of the destination address and the source address is longer than N, the rule 9 is applied. Otherwise, the order of the destination addresses do not change. The N should be configurable and it should be 32 by default. This is simply because the two sites whose matching bit length is longer than 32 are probably adjacent.

Now that IPv6 PI address is admitted in some RIRs, hierachical address assignment is not maintained anymore. It seems that the longest matching algorithm may not worth the adverse effect of disalbing the DNS based load balance technique.

After long discussion, however we could not reach any consensus here. That means, we cannot change the current rules for this issue.

B.2. NAT64 prefix issue

NAT64 WKP was newly defined[RFC6052]. It depends site by site whether NAT64 should be preferred over IPv4, in other words NAT44, or NAT44 over NAT64. So, this issue of site local policy should be solved by policy distribution mechanism.

Appendix C. Revision History

03:

ULA address selection issue was expanded.
6to4, Teredo and IPv4 priorization issue was elaborated.
Deperecated address blocks in policy table section was elaborated.
In appendix, NAT64 prefix issue was added.

02:

Suresh Krishnan's suggestions for better english sentences were incorporated.
A new source address selection rule that utilizes the next-hop information is included in Section 2.3.
Site local address prefix was corrected.

01:
   Re-structured to contain only the actual changes to RFC 3484.

00:
   Published as a 6man working group item.

03:

   Added acknowledgements.
   Added longest matching algorithm malfunction regarding local DNS round robin.
   The proposed changes section was re-structured.
   The issue of 6to4/Teredo and IPv4 prioritization was included.
   The issue of deprecated addresses was added.
   The renewed default policy table was changed accordingly.

02:
   Added the reference to address selection design team's proposal.

01:
   The issue of private IPv4 address scope was added.
   The issue of ULA address scope was added.
   Discussion of longest matching rule was expanded.

Authors' Addresses

Arifumi Matsumoto
NTT SI Lab
Midori-Cho 3-9-11
Musashino-shi, Tokyo 180-8585
Japan

Phone: +81 422 59 3334
Email: arifumi@nttv6.net

Jun-ya Kato
NTT SI Lab
Midori-Cho 3-9-11

Musashino-shi, Tokyo 180-8585
Japan

Phone: +81 422 59 2939
Email: kato@syce.net

Tomohiro Fujisaki
NTT PF Lab
Midori-Cho 3-9-11
Musashino-shi, Tokyo 180-8585
Japan

Phone: +81 422 59 7351
Email: fujisaki@syce.net