Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 4, 2013

D. Cheng
Huawei Technologies
M. Boucadair
France Telecom

July 3, 2012

Routing for IPv4-embedded IPv6 Packets
draft-ietf-ospf-ipv4-embedded-ipv6-routing-03

Abstract

This document describes routing packets destined to IPv4-embedded IPv6 addresses across IPv6 transit core using OSPFv3 with a separate routing table.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 4, 2013.

Table of Contents

1. Introduction

This document describes a routing scenario where IPv4 packets are transported over IPv6 network.

In this document the following terminology is used:

- An IPv4-embedded IPv6 address denotes an IPv6 address which contains an embedded 32-bit IPv4 address constructed according to the rules defined in [RFC6052].

- IPv4-embedded IPv6 packets are packets of which destination addresses are IPv4-embedded IPv6 addresses.

- AFBR (Address Family Border Router, [RFC5565]) refers to an edge router (PE), which supports both IPv4 and IPv6 address families, but the backbone network the PE connects to only supports IPv4 or IPv6 address family.

- AFXLBR (Address Family Translation Border Router) is defined in this document. It refers to a border router that supports both IPv4 and IPv6 address families, located on the boundary of IPv4-only network and IPv6-only network, and is capable of performing IP header translation between IPv4 and IPv6 according to [RFC6145].

1.1. The Scenario

Due to exhaustion of public IPv4 addresses, there has been continuing effort within IETF on IPv6 transitional techniques. In the course of transition, it is certain that networks based on IPv4 and IPv6 technologies respectively, will co-exist at least for some time. One scenario of the co-existence is that IPv4-only networks inter-connecting with IPv6-only networks, and in particular, when an IPv6-only network serves as a transit network that inter-connects several segregated IPv4-only networks. In this scenario, IPv4 packets are transported over the IPv6 transit network between IPv4 networks. In order to forward an IPv4 packet from a source IPv4 network to the destination IPv4 network, IPv4 reachability information must be exchanged between the IPv4 networks by some mechanisms.

In general, running an IPv6-only network would reduce OPEX and optimize the operation comparing to IPv4-IPv6 dual-stack environment. Some solutions have been proposed to allow delivery of IPv4 services over an IPv6-only network. This document focuses on an engineering techniques which aims to separate the routing table dedicated to IPv4-embedded IPv6 destination from native IPv6 ones.

Maintaining a separate routing table for IPv4-embedded IPv4 routes optimizes IPv4 packets forwarding process. It also prevents any overload of the native IPv6 routing tables. A separate routing table can be generated from a separate routing instance or a separate routing topology.

1.2. Routing Solution per RFC5565

The aforementioned scenario is described in [RFC5565], i.e.- IPv4-over-IPv6 scenario, where the network core is IPv6-only, and the inter-connected IPv4 networks are called IPv4 client networks. The P routers in the core only support IPv6 but the AFBRs (Address Family Border Routers) support IPv4 on interface facing IPv4 client networks, and IPv6 on interface facing the core. The routing solution defined in [RFC5565] for this scenario is to run i-BGP among AFBRs to exchange IPv4 routing information with each other, and the IPv4 packets are forwarded from one IPv4 client network to the other through a softwire using tunneling technology such as MPLS LSP, GRE, L2TPv3, etc.

1.3. An Alternative Routing Solution with OSPFv3

In this document, we propose an alternative routing solution for the scenario described in Section 1.1, where several segregated IPv4 networks, called IPv4 client networks, are interconnected by an IPv6 transit network. We name the border node on the boundary of an IPv4 client network and the IPv6 transit network as Address Family Translation Border Router (AFXLBR), which supports both IPv4 and IPv6 address families, and is capable of translating an IPv4 packet to an IPv6 packet, and vice versa, according to [RFC6145].

Since the scenario occurs most in a single ISP operating environment, an IPv6 prefix can be locally allocated and used to construct IPv4-embedded IPv6 addresses according to [RFC6052] by each AFXLBR. The embedded IPv4 address or prefix belongs to an IPv4 client network that is connected to the AFXLBR. An AFXLBR injects IPv4-embedded IPv6 addresses and prefixes into the IPv6 transit network using OSPFv3, and it also installs IPv4-embedded IPv6 routes advertised by other AFXLBRs.

When an AFXLBR receives an IPv4 packet from a locally connected IPv4 client network and destined to a remote IPv4 client network, it translates the IPv4 header to the relevant IPv6 header according to [RFC6145], and in that process, source and destination IPv4 address are translated into IPv4-embedded IPv6 addresses, respectively, according to [RFC6052]. The resulting IPv6 packet is then forwarded to the AFXLBR that connects to the destination IPv4 client network. The remote AFXLBR derives the IPv4 source and destination addresses

from the IPv4-embedded IPv6 addresses, respectively, according to [RFC6052], and translates the header of the received IPv6 packet to the relevant IPv4 header according to [RFC6145]. The resulting IPv4 packet is then forwarded according to the IPv4 routing table maintained on the AFXLBR.

There are use cases where the proposed routing solution is useful. One case is that some border nodes do not participate in i-BGP for routes exchange (one example is documented in [I-D.boucadair-softwire-dslite-v6only]), or i-BGP is not used at all. Another case is that tunnel mechanism is not used in the IPv6 transit network, or native IPv6 forwarding is preferred. Note that with this routing solution, the IPv4 and IPv6 header translation that occurs at an AFXLBR in both diretions is stateless.

1.4. OSPFv3 Routing with a Specific Topology

In general, IPv4-embedded IPv6 packets can be forwarded just like native IPv6 packets with OSPFv3 running in the IPv6 transit network. However, this would require IPv4-embedded IPv6 routes to be flooded throughout the entire transit network and stored on every router. This is not desirable in the scaling perspective. Moreover, since all IPv6 routes are stored in the same routing table, it is inconvenient to manage the resource required for routing and forwarding based on traffic category, if so desired.

To improve the situation, a separate OSPFv3 routing table can be constructed that is dedicated to IPv4-embedded IPv6 topology, and that table is solely used for routing IPv4-embedded IPv6 packets in the IPv6 transit network. The IPv4-embedded IPv6 topology include all the participating AFXLBR routers and a set of P routers for connectivity and routing paths.

There are two methods to build a separate OSPFv3 routing table for IPv4-embedded IPv6 routes as follows:

- The first one is to run a separate OSPFv3 instance for IPv4-embedded IPv6 topology in the IPv6 transit network according to [RFC5838].

- The second one is to stay with the existing OSPFv3 instance that already operates in the transit network, but maintain a separate IPv4-embedded IPv6 topology, according to [I-D.ietf-ospf-mt-ospfv3].

With either method, there would be a dedicated IPv4-embedded IPv6 topology that is maintained on all participating AFXLBR and P routers, along with a dedicated IPv4-embedded IPv6 routing table.

The routing table is then used solely in the IPv6 transit network for IPv4-embedded IPv6 packets.

It would be an operator's preference as which method is to be used. This document elaborates on how configuration is done for each method and related routing issues that is common to both.

This document only focuses on unicast routing for IPv4-embedded IPv6 packets using OSPFv3.

2. Provisioning

2.1. Deciding the IPv4-embedded IPv6 Topology

Before making appropriate configuration in order to generate a separate OSPFv3 routing table for IPv4-embedded IPv6 addresses and prefixes, decision must be made on the set of routers and their interfaces in the IPv6 transit network that should be on the IPv4-embedded IPv6 topology.

For the purpose of this topology, all AFXLBRs that connect to IPv4 client networks must be members of this topology, and also at least some of their network core facing interfaces along with some P routers in the IPv6 transit network would be on this topology.

The IPv4-embedded IPv6 topology is a sub-topology of the entire IPv6 transit network, and if all routers (including AFXLBRs and P-routers) and all their interfaces are included, the two topologies converge. In general, as more P routers and their interfaces are configured on this sub-topology, it would increase the inter-connectivity and potentially, there would be more routing paths across the transit network from one IPv4 client network to the other, at the cost that more routers need to participate the IPv4-embedded IPv6 routing. In any case, the IPv4-embedded IPv6 topology must be continuous with no partitions.

2.2. Maintaining a Dedicated IPv4-embedded IPv6 Routing Table

In an IPv6 transit network, in order to maintain a separate IPv6 routing table that contains routes for IPv4-embedded IPv6 destinations only, OSPFv3 needs to use the mechanism defined either in [RFC5838] or in [I-D.ietf-ospf-mt-ospfv3] with required configuration, as described in the following sub-sections.

2.3. OSPFv3 Topology with a Separate Instance ID

It is assumed that the scenario as described in this document is under a single ISP and as such, an OSPFv3 instance ID (IID) is allocated locally and used for an OSPFv3 operation dedicated to unicast IPv4-embedded IPv6 routing in an IPv6 transit network. This IID is configured on each OSPFv3 interface of routers that participates in this routing instance.

The range for a locally configured OSPFv3 IID is from 128 to 255, inclusively, and this number must be used to encode the "Instance ID" field in the OSPFv3 packet header on every router that executes this instance in the IPv6 transit network.

In addition, the "AF" bit in the OSPFv3 Option field must be set.

During the Hello packets processing, adjacency may only be established when received Hello packets contain the same Instance ID as configured on the receiving interface for OSPFv3 instance dedicated to the IPv4-embedded IPv6 routing.

For more details, the reader is referred to [RFC5838].

2.4. OSPFv3 Topology with the Default Instance

Similar to that as described in the previous section, an OSPFv3 multi-topology ID (MT-ID) is locally allocated and used for an OSPFv3 operation including unicast IPv4-embedded IPv6 routing in an IPv6 transit network. This MTID is configured on each OSPFv3 interface of routers that participates in this routing topology.

The range for a locally configured OSPFv3 MT-ID is from 32 to 255, inclusively, and this number must be used to encode the "MT-ID" field that is included in some of the extended LSAs as documented in [I-D.ietf-ospf-mt-ospfv3].

In addition, the MT bit in the OSPFv3 Option field must be set.

For more details, the reader is referred to [I-D.ietf-ospf-mt-ospfv3].

3. IP Packets Translation

When transporting IPv4 packets across an IPv6 transit network with the mechanism described above, an IPv4 packet is translated to an IPv6 packet at ingress AFXLBR, and the IPv6 packet is translated back to the original IPv4 packet at egress AFXLBR. The IP packet

translation is accomplished in stateless manner according to rules specified in [RFC6145], with the address translation detail explained in the next sub-section.

3.1. Address Translation

Prior to the operation, an IPv6 prefix is allocated by the ISP and it is used to form IPv4-embedded IPv6 addresses.

The IPv6 prefix can either be a well-known IPv6 prefix (WKP) 64: ff9b::/96, or a network-specific prefix that is unique to the ISP; and for the later case, the IPv6 prefix length may be 32, 40, 48, 56 or 64. In either case, this IPv6 prefix is used during the address translation between an IPv4 address and an IPv4-embedded IPv6 address, which is performed according to [RFC6052].

During translation from an IPv4 header to an IPv6 header at an ingress AFXLBR, the source IPv4 address and destination IPv4 address are translated into the corresponding IPv6 source address and destination IPv6 address, respectively, and during translation from an IPv6 header to an IPv4 header at an egress AFXLBR, the source IPv6 address and destination IPv6 address are translated into the corresponding IPv4 source address and destination IPv4 address, respectively. Note that the address translation is accomplished in a stateless manner.

4. Advertising IPv4-embedded IPv6 Routes

In order to forward IPv4 packets to the proper destination across IPv6 transit network, IPv4 reachability needs to be disseminated throughout the IPv6 transit network and this work is performed by AFXLBRs that connect to IPv4 client networks using OSPFv3.

With the scenario described in this document, i.e. - a set of AFXLBRs that inter-connect a bunch of IPv4 client networks with an IPv6 transit network, we view that IPv4 networks and IPv6 networks belong to separate Autonomous Systems, and as such, these AFXLBRs are OSPFv3 ASBRs.

4.1. Advertising IPv4-embedded IPv6 Routes into IPv6 Transit Network

IPv4 addresses and prefixes in an IPv4 client network are translated into IPv4-embedded IPv6 addresses and prefixes, respectively, using the same IPv6 prefix allocated by the ISP and the method specified in [RFC6052]. These routes are then advertised by one or more attached ASBRs into the IPv6 transit network using AS External LSA [RFC5340], i.e. - with the advertising scope throughout the entire Autonomous

System.

4.1.1. Routing Metrics

By default, the metric in an AS External LSA that carries an IPv4-embedded IPv6 address or prefixes is a Type 1 external metric, which is then to be added to the metric of an intra-AS path during OSPFv3 routes calculation. By configuration on an ASBR, the metric can be set to a Type 2 external metric, which is considered much larger than that on any intra-AS path. The detail is referred to OSPFv3 specification [RFC5340]. In either case, an external metric may take the same value as in an IPv4 network (running OSPFv2 or others), but may also be specified based on some routing policy; the detail is outside of the scope of this document.

4.1.2. Forwarding Address

If the "Forwarding Address" field of an OSPFv3 AS External LSA is used to carry an IPv6 address, that must also be an IPv4-embedded IPv6 address where the embedded IPv4 address is the destination address in an IPv4 client network. However, since an AFXLBR sits on the border of an IPv4 network and an IPv6 network, it is recommended that the "Forwarding Address" field not to be used by setting the F bit in the associated OSPFv3 AS-external-LSA to zero, so that the AFXLBR can make the forwarding decision based on its own IPv4 routing table.

4.2. Advertising IPv4 Addresses into Client Networks

IPv4-embedded IPv6 routes injected into the IPv6 transit network from one IPv4 client network may be advertised into another IPv4 client network, after the associated destination addresses and prefixes are translated back to IPv4 addresses and prefixes, respectively. This operation is similar to the regular OSPFv3 operation, wherein an AS External LSA can be advertised in a non-backbone area by default.

An IPv4 client network that does not want to receive such advertisement can be configured as a stub area or with other routing policy.

For the purpose of this document, IPv4-embedded IPv6 routes must not be advertised into any IPv6 client networks that also connected to the IPv6 transit network.

5. Aggregation on IPv4 Addresses and Prefixes

In order to reduce the amount of AS External LSAs that are injected

to the IPv6 transit network, effort must be made to aggregate IPv4 addresses and prefixes at each AFXLBR before advertising.

6. Forwarding

There are three cases in forwarding IP packets in the scenario as described in this document, as follows:

1. On an AFXLBR, if an IPv4 packet that is received on an interface connecting to an IPv4 client network with the destination IPv4 address belong to another IPv4 client network, the header of the packet is translated to a corresponding IPv6 header as described in Section 3, and the packet is then forwarded to the destination AFXLBR that advertises the IPv4-embedded IPv6 address through the IPv6 transit network.

2. On an AFXLBR, if an IPv4-embedded IPv6 packet is received and the embedded destination IPv4 address is in its IPv4 routing table, the header of the packet is translated to a corresponding IPv4 header as described in Section 3, and the packet is then forwarded accordingly.

3. On any router that is within the IPv4-embedded IPv6 topology located in the IPv6 transit network, if an IPv4-embedded IPv6 packet is received and a route is found in the IPv4-embedded IPv6 routing table, the packet is forwarded accordingly.

The classification of IPv4-embedded IPv6 packet is according to the IPv6 prefix of the destination address, which is either the Well Known Prefix (i.e., 64:ff9b::/96) or locally allocated as defined in [RFC6052].

7. MTU Issues

In the IPv6 transit network, there is no new MTU issue introduced by this document. If a separate OSPFv3 instance (per [RFC5838]) is used for IPv4-embedded IPv6 routing, the MTU handling in the transit network is the same as that of the default OSPFv3 instance. If a separate OSPFv3 topology (according to [I-D.ietf-ospf-mt-ospfv3]) is used for IPv4-embedded IPv6 routing, the MTU handling in the transit network is the same as that of the default OSPFv3 topology.

However, the MTU in the IPv6 transit network may be different than that of IPv4 client networks. Since an IPv6 router will never fragment a packet, the packet size of any IPv4-embedded IPv6 packet entering the IPv6 transit network must be equal to or less than the

MTU of the IPv6 transit network. In order to achieve this requirement, it is recommended that AFXLBRs to perform IPv6 path discovery among themselves and the resulting MTU, after taking into account of the difference between IPv4 header length and IPv6 header length, must be "propagated" into IPv4 client networks, e.g.- included in the OSPFv2 Database Description packet.

The detail of passing the proper MTU into IPv4 client networks is beyond the scope of this document.

8. Backdoor Connections

In some deployments, IPv4 client networks are inter-connected across the IPv6 transit network, but also directly connected to each other. The "backdoor" connections between IPv4 client networks can certainly be used to transport IPv4 packets between IPv4 client networks. In general, backdoor connections are prefered over the transportation over the IPv6 transit network, since there requires no address family translation.

9. Security Considerations

This document does not introduce any security issue than what has been identified in [RFC5838] and [RFC6052].

10. IANA Considerations

No new IANA assignments are required for this document.

11. Acknowledgements

Many thanks to Acee Lindem, Dan Wing and Joel Halpern for their comments.

12. References

12.1. Normative References

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.

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

October 2010.

[RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.

12.2. Informative References

[I-D.boucadair-softwire-dslite-v6only]
Boucadair, M., Jacquenet, C., Grimault, J., Kassi-Lahlou, M., Levis, P., Cheng, D., and Y. Lee, "Deploying Dual-Stack Lite in IPv6 Network", draft-boucadair-softwire-dslite-v6only-01 (work in progress), April 2011.

[I-D.ietf-ospf-mt-ospfv3]
Mirtorabi, S. and A. Roy, "Multi-topology routing in OSPFv3 (MT-OSPFv3)", draft-ietf-ospf-mt-ospfv3-03 (work in progress), July 2007.

[RFC5565] Wu, J., Cui, Y., Metz, C., and E. Rosen, "Softwire Mesh Framework", RFC 5565, June 2009.

[RFC5838] Lindem, A., Mirtorabi, S., Roy, A., Barnes, M., and R. Aggarwal, "Support of Address Families in OSPFv3", RFC 5838, April 2010.

Authors' Addresses

Dean Cheng
Huawei Technologies
2330 Central Expressway
Santa Clara, California 95050
USA

Email: dean.cheng@huawei.com

Mohamed Boucadair
France Telecom
Rennes, 35000
France

Email: mohamed.boucadair@orange.com