Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 8, 2011

M. Boucadair, Ed.
C. Jacquenet
JL. Grimault
M. Kassi-Lahlou
P. Levis
France Telecom
D. Cheng, Ed.
Huawei Technologies Co., Ltd.
Y. Lee, Ed.
Comcast
October 5, 2010

Deploying Dual-Stack Lite in IPv6 Network
draft-boucadair-softwire-dslite-v6only-00

Abstract

Dual-Stack lite requires that the AFTR must have IPv4 connectivity. This forbids a service provider who wants to deploy AFTR in an IPv6-only network. This memo proposes an extension to implement a stateless IPv4-in-IPv6 encapsulation in the AFTR so that AFTR can be deployed in an IPv6-only network.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

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 April 8, 2011.

Copyright Notice

Table of Contents

1. Introduction 4
1.1. General 4
1.2. Requirements 4
1.3. Overview 4
2. Terminology 6
3. Addressing 7
4. DS-Lite AFTR 8
4.1. Provisioning 8
4.2. Procedure 8
4.2.1. Processing Ingress Traffic from Customer Interface . . 9
4.2.2. Processing Ingress Traffic from Core Interface 9
4.3. Flows Examples 10
5. IPv6-IPv4 Interconnection Function (ICXF) 11
5.1. Provisioning 11
5.2. Procedure 11
6. Routing Architecture and Considerations 12
6.1. Static Routing 12
6.2. Dynamic Routing 12
7. Multicast Considerations 14
8. Fragmentation 14
9. Conclusions 14
10. IANA Considerations 15
11. Security Considerations 15
12. Acknowledgements 15
13. References 15
13.1. Normative References 15
13.2. Informative References 16
Appendix A. Changes Since 02 16
Authors' Addresses 16

1. Introduction

1.1. General

Dual-Stack lite (DS-lite) contains two major concepts: (1) Transport IPv4 packets over an IPv6 access network, and (2) Share a public IPv4 address to multiple users.

The B4 element resided in the customer premises is provisioned an global routable IPv6 address. It also establishes an IPv4-in-IPv6 tunnel to AFTR element. The hosts behind B4 elements are assigned with [RFC1918] addresses. When the B4 receives IPv4 datagram from it managed host, it will send the datagram over the IPv4-in-IPv6 tunnel to the AFTR.

AFTR element provides the NAT function and is responsible for sharing public IPv4 addresses to multiple B4 elements. It also requires direct IPv4 connectivity to send and received the NAT-ed datagram to the IPv4 network.

This model puts a demarcation in the network where the access network between B4 and AFTR can be IPv6-only and the network north of AFTR must be IPv4. Consider a service provider wants to extend the IPv6-only network boundary from the access network to the border of the network, this will force the AFTR to be deployed in the border and further away from the B4s. This memo describes a framework to allow a service provider to extend the IPv6-only network while to allow the AFTR to stay close to the B4s.

1.2. Requirements

- [REQ1] Extend the IPv6-only boundary to the border of the network.
- [REQ2] Only the AS Border Router has IPv4 connectivity.
- [REQ3] The service provider provisions only IPv6 addresses to the customers but continues to provide IPv4 services to them.
- [REQ4] The AFTR has only IPv6-connectivity and must be able to send and receive IPv4 packets.

1.3. Overview

DS-Lite [I-D.ietf-softwire-dual-stack-lite] directly connects users to IPv6 networks but at the same time provides IPv4 services by tunneling IPv4 packets over an IPv6 network.

AFTR element is the combination of an IPv4-in-IPv6 tunnel end-point

and an IPv4-IPv4 NAT implemented in the same node. In addition, the specification assumes that an AFTR is directly connected to the IPv4 network.

In some deployments where the service provider wants to deploy AFTR in the IPv6 core network. AFTR nodes may not have direct IPv4 connectivity. In this scenario, IPv4 packets after NAT44 function applied on an AFTR node need to be transported over the IPv6 core network to the IPv4 network. This memo proposes a framework for this scenario as an extension to the DS-Lite specification.

In this specification, we define a new stateless IPv6-IPv4 interconnection function (referred to as IPv6-IPv4 ICXF), in a border node located at the boundaries between the IPv6 and IPv4 networks. The AFTR discovers the ICXF address, and sends IPv4 encapsulated IPv6 packets after NAT44 function.

The ICXF may be hosted in an ASBR (Autonomous System Border Router) or a dedicated node located at the interconnection between IPv6 and IPv4 domains. A router that hosts the ICXF is referred to as an ICXF router.

When the AFTR receives a customer’s outbound packet from B4 element, it de-capsulate the packet and perform standard NAT44 function. Since an AFTR does not directly connect to the IPv4 network, AFTR will encapsulate the NAT-ed IPv4 packet in an IPv4-in-IPv6 packet, with an IPv4-Embedded IPv6 destination address [I-D.ietf-behave-address-format], and forward it to an ICXF router located with direct connection to the IPv4 network. When the ICXF router receives the IPv4-Embedded IPv6 packet, it will de-capsulate the packet and forward the IPv4 packet based on the IPv4 destination address.

For an inbound IPv4 packet to B4 element, the ICXF router will encapsulate the IPv4 packet into IPv6 packet with the IPv4-Embedded IPv6 address and forward it to the appropriate DS-Lite AFTR node, which de-capsulates the IPv4 packet and then follows the normal procedure defined by DS-Lite architecture as if the IPv4 packet is received from a directly connected IPv4 network.

Figure 1 provides an overview of the global architecture. Customers are connected to the service domain via a CPE device. Several DS-Lite AFTR nodes are deployed to manage the traffic sent and received by the end-user terminal devices. The service domain is IPv6 only and interconnection with adjacent IPv4 realms is implemented using IPv6-IPv4 ICXF. The distributed deployment mode of AFTR nodes is motivated by several reasons such as optimizing intra-domain paths, avoiding single point of failure, minimizing the impact on geo-

localization services, minimizing the amount of customers to be impacted by an AFTR node failure, etc. AFTR deployment model varies from service provider to service provider and it is out of scope of this specification.

Note in this architecture, the DS-Lite B4 element (located in a CPE) and AFTR still behave exactly as defined in [I-D.ietf-softwire-dual-stack-lite], but with additional functions added to the AFTR when it does not directly connect to the IPv4 network. A new ICXF function is introduced to perform stateless IPv6- IPv4 interconnection. This specification defines new requirements on addressing scheme and routing. More details are provided in the following sections.

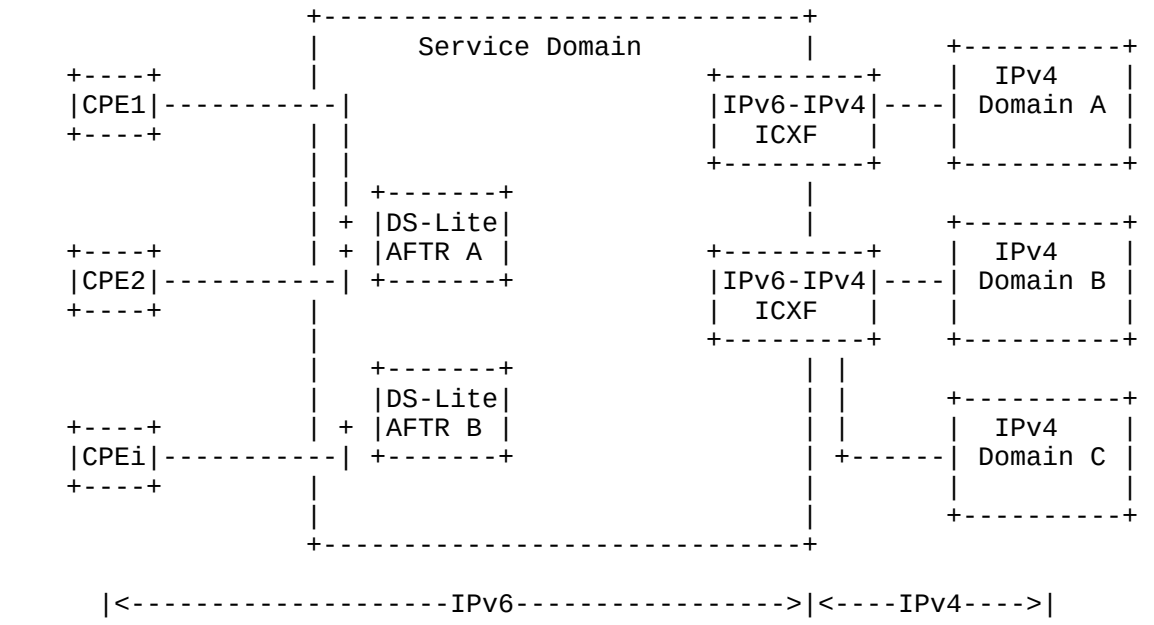


Figure 1: Architecture Overview

2. Terminology

This memo defines the following terms:

- IPv6-IPv4 Interconnection Function (IPv6-IPv4 ICXF): refers to the function that de-capsulates (resp., encapsulates) the IPv6 (resp.,

IPv4) packet from DS-Lite AFTR node(s) and forwards the IPv4 (resp., IPv6) packets to the IPv4 (resp., IPv6) networks.

- ICXF router: refers to the border router implemented with IPv6-IPv4 ICXF.

- DS-Lite AFTR node: refers to the AFTR node whose behavior is specified in [I-D.ietf-softwire-dual-stack-lite]. In addition, this specification assumes that the DS-Lite AFTR node is only connected to an IPv6 network. The AFTR will encapsulate the IPv4 packet in an IPv6 packet (IPv4-in-IPv6) after the NAT44 function. The encapsulated IPv6 packet will be forwarded to the ICXF router. This IPv4-inIPv6 encapsulation is stateless.

- Access segment: This segment encompasses both the IP access to the customers and to the service provider's network.

- Interconnection segment: Includes all nodes and resources which are deployed at the border of a given AS (Autonomous System) a la BGP.

- Core segment: Denotes a set of IP networking capabilities and resources which are located between the interconnection and the access segments.

- Pref6: An IPv6 prefix assigned by LIR. This prefix is configured on both ICXF and AFTR.

- FROM-AFTR Address: An IPv4-Embedded IPv6 address [I-D.ietf-behave-address-format] that combines an IPv6 prefix Pref6 and the destination IPv4 address.

- TO-AFTR Address: An IPv4-Embedded IPv6 address [I-D.ietf-behave-address-format] that combines an IPv6 prefix Pref6 and a destination IPv4 address which configured in an AFTR NAT pool.

3. Addressing

For outbound IPv4 packets, the AFTR performs encapsulation and the ICXF router performs de-capsulation. For inbound IPv4 packets, the ICXF router performs IPv4-in-IPv6 encapsulation and an AFTR performs de-capsulation.

When an AFTR forwards an IPv6 packet with an IPv4 payload to an ICXF router, the source IPv6 address is one of the AFTR's IPv6 address, which is normally a global IPv6 address configured on an interface of

the node (e.g., an address of a loopback interface), and the destination IPv6 address is the FROM-AFTR Address.

When an ICXF router receives an IPv4 packet, it encapsulates the IPv4 packet with an IPv6 header where the source IPv6 address is the ICXF router's global IPv6 address and the destination IPv6 address is the TO-AFTR address. The TO-AFTER address is constructed by combining the Pref6 and the destination IPv4 address in the IPv4 packet. The destination IPv4 address is one of the addresses configured in the AFTR NAT pool.

In both cases, the Pref6 is an IPv6 prefix assigned by the service provider, and is used to construct an IPv4-Embedded IPv6 address. Figure 2 gives an example of the address format.

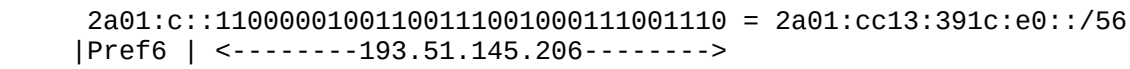


Figure 2: Example for an IPv4-Embedded IPv6 Prefix

In this example, Pref6 is 2a01:c::/20 and the IPv4_Addr is 193.51.145.206. Then, the corresponding IPv6 prefix is: 2a01:cc13:391c:e0::/56. We use a /20 prefix for Pref6. However, an operator can decide to use any prefix length.

4. DS-Lite AFTR

4.1. Provisioning

The AFTR must be provisioned with a set of global IPv4 prefixes for NAT44 operations. In addition, an IPv6 prefix (i.e., Pref6) is configured in the AFTR. The Pref6 is used to construct FROM-AFTR addresses. The FROM-AFTR addresses are used in the destination address field of the IPv6 header for the IPv4-in-IPv6 packets.

4.2. Procedure

Figure 3 shows the input and output of a DS-Lite AFTR node.

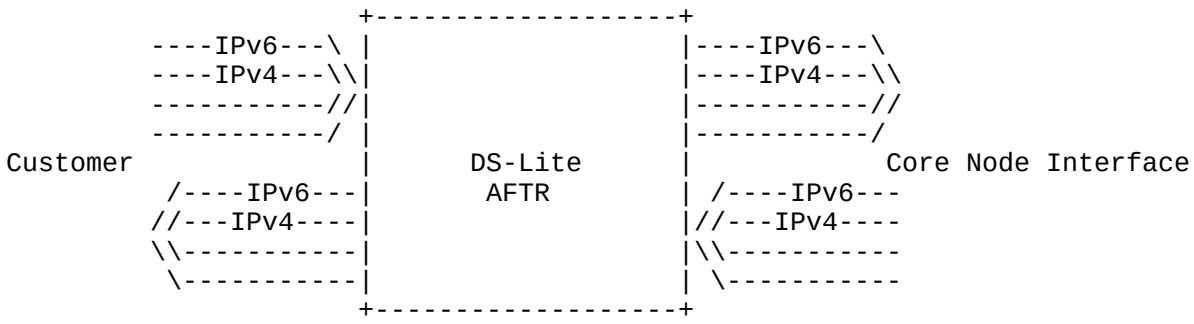


Figure 3: Modified DS-Lite AFTR

Two main (logical) interfaces may be distinguished in a DS-Lite AFTR node as follows:

a. Interface with the customer device, i.e.- DS-Lite interface per [I-D.ietf-softwire-dual-stack-lite].

b. Interface with core nodes. Note that the DS-Lite AFTR does not directly connect to an IPv4 domain.

4.2.1. Processing Ingress Traffic from Customer Interface

1. De-capsulate the IPv6 header from the IPv4-in-IPv6 packets (sent by the customer device) 2.

2. NAT the IPv4 packet and create an entry in the NAT binding table

3. Encapsulate the NAT-ed IPv4 packets in IPv6 with a destination IPv6 address built according to the addressing scheme described in Section 3. Encapsulated packet is forwarded based on the FROM-AFTR IPv6 address by standard routing. Depending on the target IPv4 address, the destination can be an AFTR node inside the service provider’s domain if the IPv4 address is one of the addresses owned by another AFTR (See Figure 4). Or, the destination can be an ICXF router if the IPv4 address is external to the service provider.

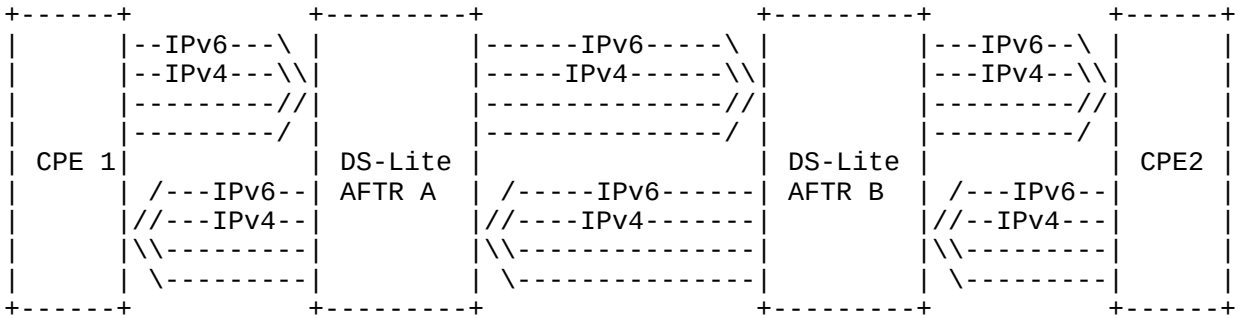
4.2.2. Processing Ingress Traffic from Core Interface

1. De-capsulate the IPv6 header and extract the IPv4 packet.

2. Process the embedded IPv4 packet according to [I-D.ietf-softwire-dual-stack-lite].

3. Forward the resulting IPv6 packet to the corresponding B4.

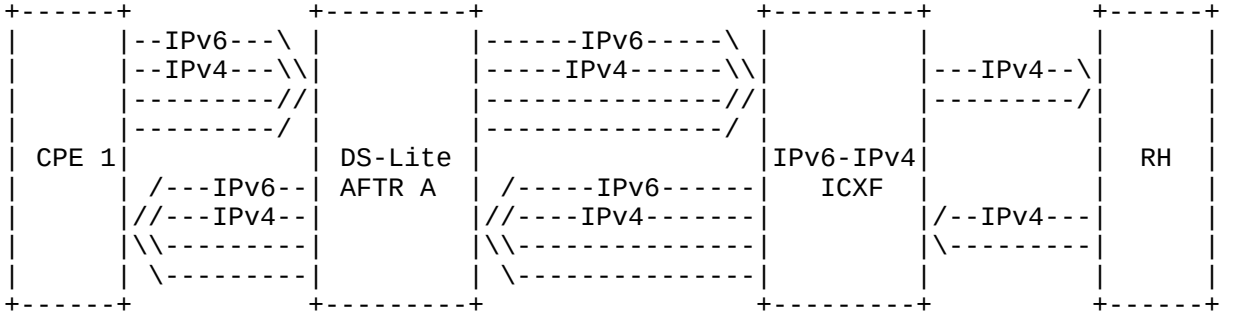
4.3. Flows Examples



Note that hosts connected to each CPE are not represented in the figure.

Figure 4: Flow Example involving two devices attached to distinct AFTRs

The following figure illustrates an example of CPE connected to a DS-Lite AFTR node, which establishes a communication with a remote host (referred to as RH) which is on an IPv4 network.



Note that host connected to CPE1 are not represented in the figure.

Figure 5: Flow Example involving only one device attached to a DS-lite enabled domain

5. IPv6-IPv4 Interconnection Function (ICXF)

ICXF is a border element that encapsulate IPv4 packets from external IPv4 network to AFTR and de-capsulate IPv6 packets from AFTR to external IPv4 network

Externally, the ICXF is connected to IPv4 network. It is an IPv4 router and performs standard IPv4 routing. It contains an IPv4 routing table and exchanges IPv4 prefixes to the internal and external peers.

Internally, the ICXF is connected to an IPv6 network and exchanges IPv4 prefixes to the AFTRs. Section 6 discusses the internal routing in details.

5.1. Provisioning

An IPv6-IPv4 ICXF router is provisioned with an IPv6 prefix (i.e., Pref6). Pref6 is used to build TO-AFTR addresses.

5.2. Procedure

Figure 6 shows the input and output of an IPv6-IPv4 ICXF.

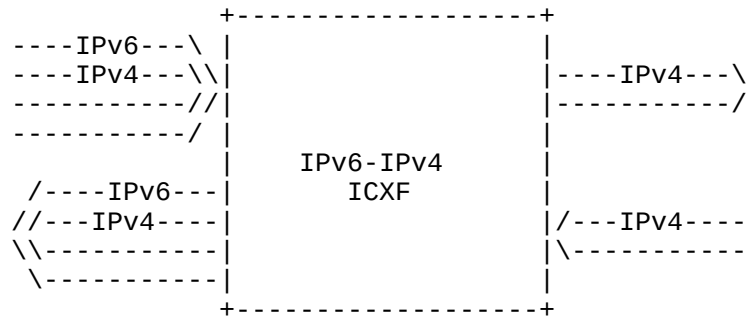


Figure 6: IPv6-IPv4 Interworking Function

When the ICXF router receives an IPv4 packet from an external IPv4 domain, it encapsulates the IPv4 packet in IPv6 packet using the following information:

- Source IPv6 address: One of its own IPv6 addresses.
- Destination IPv6 address: TO-AFTR Address which is an IPv4-Embedded IPv6 address using the Pref6 and destination IPv4 address

of the encapsulated IPv4 packet.

As for the outbound IPv6 packets, an ICXF router performs de-capsulation and forwards the embedded IPv4 packets to the connected IPv4 networks according to IPv4 routing rule.

6. Routing Architecture and Considerations

This section describes the routing consideration to support this specification, i.e.- how an IPv6 packet with an IPv4-Embedded IPv6 destination address is forwarded from a DS-Lite AFTR to an ICXF router, and vice versa, in the IPv6 network.

When a DS-Lite AFTR forwards IPv4-in-IPv6 packets to an ICXF router, the destination IPv6 address is an IPv4-Embedded IPv6 address, where the Pref6 is an IPv6 prefix assigned to the service provider and the IPv4 address is reachable through one or more ICXF routers. The forwarding decision can be made based on static or dynamic routing.

6.1. Static Routing

The AFTR is configured with static routes, and each static route points to an IPv4-Embedded IPv6 prefix. Alternatively, the AFTR can contain a default route where the default is the ICXF.

6.2. Dynamic Routing

Dynamic routing is more desirable for the deployments where there are multiple DS-Lite AFTRs and ICXF routers. This specification suggests four dynamic routing options as documented below:

Option 1:

- AFTRs and ICXF routers are configured as a Softwire Mesh [RFC5565] and iBGP is used to exchange IPv4 reachability information. AFTR and ICXF will peer with each other over iBGP and exchange their IPv4 reachability. Each AFTR and ICXF will compute an IPv4 routing table based upon the BGP table. Given an IPv4 network managed by the AFTR or ICXF, the next-hop of this network is the IPv6 address of the AFTR or ICXF.

- Pros: This routing option offers an optimized forwarding for IPv4-in-IPv6 packets in the IPv6 network.

- Cons: DS-Lite AFTRs and ICXF routers must peer in iBGP and storing BGP routes on all these nodes.

Option 2:

- ICXF router advertises its IPv4 reachability information in IGP. This routing option does not require AFTRs and ICXFs to be iBGP peers. For the AFTR IPv6 routing table, it contains all FROM_AFTR prefixes and the ICXF IPv4 reachability information in the form on IPv4-Embedded IPv6 prefixes (i.e., Pref6 + ICXF IPv4 routing information).

- Pros: Given that the ICXF advertises all its IPv4 network reachability in IPv6 network, the AFTR can choose the best path to forward the packet.

- Cons: This optimization has a drawback: ICXF routers are required to advertise its full IPv4 reachability to in IGP. As such, IPv6 routers will maintain the full IPv4 reachability in its IPv6 routing table.

Option 3:

- With this option, each ICXF router advertises a Pref6 (Section 5.1) in the IPv6 IGP. An AFTR forwards an IPv4-in-IPv6 packet always to a nearest ICXF router. In other words, the nearest ICXF is the default router for all external IPv4 prefixes.

- Pros: Significantly reduces the size of the IPv6 routing table to virtually one entry for IPv4 reachability.

- Cons: The closest ICXF router may not have the best route to the final destination in the IPv4 network. The ICXF may forward the packet to another ICXF to reach the IPv4 destination based upon the local IPv4 routing information.

Option 4:

- This option requires every router in the IPv6 network to learn the IPv4-Embedded IPv6 prefixes advertised by the AFTR and ICXF. These prefixes are only meaningful to the AFTR and ICXF, other IPv6 routers are not interested in them. To address this issue, a new topology [RFC4915] or [RFC5120] can be created to store the IPv4-Embedded IPv6 prefixes.

- This option requires the ICXF router and AFTR advertise the IPv4-Embedded IPv6 prefixes in the IPv4-Embedded IPv6 topology. This topology contains only the IPv4-Embedded IPv6 prefixes. Regular IPv6 routers will not participate this topology.

- With this option, each ICXF router advertises its reachable IPv4 prefixes in the form of the IPv4-Embedded IPv6 addresses. These LSAs will appear only in the dedicated MT. AFTR which participates the MT will install the LSAs to its IPv6 routing table. Those didn't participates the MT will simply ignore the LSAs.

- Pros: Only the AFTR and ICXF install the IPv4-Embedded IPv6 prefixes in the IPv6 routing table.

- Cons: Addition administration cost to maintain a new topology in ICXF and AFTR.

7. Multicast Considerations

This document describes an IPv4-IPv6 inter-connection extension to DS-Lite [I-D.ietf-softwire-dual-stack-lite], which currently limits the scope to transporting unicast IPv4 traffic over IPv6 network only. Considerations on transporting multicast IPv4 traffic over IPv6 network is out of scope.

8. Fragmentation

Tunneling IPv4 over IPv6 between AFTR and ICXF reduce the effective MTU size by the size of an IPv6 header. Since ICXF tunnel is stateless, the tunnel endpoint can't fragment and re-assumable the oversized IPv4 packet. A service provider may increase the MTU size by 40-bytes on the IPv6 network. If this is not possible, AFTR and ICXF may use IPv6 Path MTU discovery.

ICXF nodes are stateless and not necessary to implement IPv4 fragmentation.

9. Conclusions

This document describes the mechanism to enable AFTR to operate on an IPv6-only network while offering:

- Global IPv6 <==> IPv6 communications.

- Global IPv4 <==> IPv4 communications.

- A remote IPv6 host would reach a host connected to the DS-Lite enabled domain using IPv6.

- o A remote IPv4 host would reach a host connected to the DS-Lite enabled domain using IPv4-in-IPv6.

10. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

11. Security Considerations

Security considerations defined in [I-D.ietf-softwire-dual-stack-lite] should be taken into account. In addition, current interconnection practices for ingress traffic filtering should be enforced in the interconnection points (ICXF).

12. Acknowledgements

The authors would like to thank Eric Burgey for his support and suggestions.

13. References

13.1. Normative References

[I-D.ietf-softwire-dual-stack-lite] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", draft-ietf-softwire-dual-stack-lite-06 (work in progress), August 2010.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

13.2. Informative References

[I-D.ietf-behave-address-format]
Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", draft-ietf-behave-address-format-10 (work in progress), August 2010.

[RFC4277] McPherson, D. and K. Patel, "Experience with the BGP-4 Protocol", RFC 4277, January 2006.

[RFC4915] Psenak, P., Mirtorabi, S., Roy, A., Nguyen, L., and P. Pillay-Esnault, "Multi-Topology (MT) Routing in OSPF", RFC 4915, June 2007.

[RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", RFC 5120, February 2008.

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

Appendix A. Changes Since 02

The following changes have been made since the last version:

1. Add a new section to define addressing scheme;

2. Add a new section to list all routing options in the IPv6 network;

3. Various editorial changes.

Authors' Addresses

Mohamed Boucadair (editor)
France Telecom
3, Av Francois Chateaux
Rennes 35000
France

Email: mohamed.boucadair@orange-ftgroup.com

Christian Jacquenet
France Telecom

Email: christian.jacquenet@orange-ftgroup.com

Jean-Luc Grimault
France Telecom
France

Email: jeanluc.grimault@orange-ftgroup.com

Mohammed Kassi-Lahlou
France Telecom

Email: mohamed.kassilahlou@orange-ftgroup.com

Pierre Levis
France Telecom

Email: pierre.levis@orange-ftgroup.com

Dean Cheng (editor)
Huawei Technologies Co., Ltd.

Email: Chengd@huawei.com
URI: http://www.huawei.com

Yiu L. Lee (editor)
Comcast

Email: yiu_lee@cable.comcast.com
URI: http://www.comcast.com