Network Working Group
Internet-Draft
Intended status: Informational
Expires: February 1, 2013

I. Farrer
Deutsche Telekom AG
A. Durand
Juniper Networks

July 31, 2012

lw4over6 Deterministic Architecture
draft-farrer-softwire-lw4o6-deterministic-arch-00

Abstract

This memo provides an operational deterministic architecture for the deployment of Lightweight 4over6 [I-D.cui-softwire-b4-translated-ds-lite] offering scalability and high-availability whilst preserving the per-flow stateless nature of the solution.

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 [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 February 1, 2013.

Copyright Notice

Table of Contents

1. Introduction

DS-Lite [RFC6333] is a solution to deal with the IPv4 exhaustion problem once an IPv6 access network is deployed. It enables unmodified IPv4 applications to access the IPv4 Internet over the IPv6 access network. In the DS-Lite architecture, global IPv4 addresses are shared amongst subscribers as the concentrator (AFTR) performs a Carrier-Grade NAT (CGN) function.

[I-D.cui-softwire-b4-translated-ds-lite] extends the original DS-Lite model so that NAT is performed by the CPE and IPv4 address sharing is possible through the use of source port-restrictions.

This memo provides an operational architecture for the deployment of Lightweight 4over6 [I-D.cui-softwire-b4-translated-ds-lite] offering scalability and high-availability whilst preserving the per-flow stateless nature of the solution.

The approach presented here is stateless and deterministic. It leverages the stateless properties of Lightweight 4over6 to offer a completely deterministic solution. The bindings between IPv4 addresses, ports and IPv6 addresses are pre-computed and stored identically in the AFTRs and DHCP servers. This allows for a very simple fail-over mechanism within a cluster of identically provisioned AFTRs.

2. Deterministic Architecture

2.1. Distribution of the Subscriber Population

In a large deployment, it makes sense to distribute the subscriber population into subscriber groups, managed by a single AFTR, or by many AFTRs grouped in a cluster. Topological considerations and geographical proximity may also be factors in the grouping of subscribers. The exact size of those groups depend on the capacity

characteristics of the AFTRs and is out of scope for this memo.

Each subscriber group is assigned an IPv6 anycast address and a pool of IPv4 addresses which are common to all AFTRs in a cluster. The IPv4 pool must be sized to handle the subscriber population. No constraints are placed upon the addresses that are used for this pool, in that they can be taken from a single, contiguous block, multiple non-contiguous blocks or single IPv4 addresses as required by the operator.

The exact ratio subscribers to IPv4 addresses, (e.g. the average number of ports assigned per subscriber) is out of scope for this

memo.

2.2. AFTR Cluster

All AFTRs within a cluster are configured with identical lW4o6 parameters. In particular, they are configured with the same:

- IPv6 AFTR tunnel end-point address
- IPv4 public pool
- IPv6 address to IPv4 address and port binding table

2.3. IPv4 Address Plus Ports to IPv6 Binding Table Considerations

The DHCPv4 over IPv6 server will provide each IPv6 CPE an IPv4 address and port range to use within its local NAT binding table. The DHCPv4 server uses the IPv6 address of the CPE as its identifier. As such, the DHCPv4 server contains a table for assigning a specific IPv4 address and ports based on the IPv6 address of the requesting CPE. To maintain the stateless nature of the architecture, DHCPv4 reservation based address assignment is recommended. The lease time for the IPv4 address is unimportant, although a long lease time (or even infinity) is recommended to reduce the number of DHCPv4 requests.

A similar table (containing the same address/port binding information) is also present on all AFTRs in the cluster.

The following table shows sample CPE configuration data for a subscriber group. In order for the system to function coherently, this data needs to be kept synchronised between all of the functional elements (AFTRs and DHCPv4o6 servers) serving the subscriber group.

IPv6 address	IPv4 address	port-range
2001:db8::1	1.2.3.4	1000-1999
2001:0:1::2	1.2.3.4	2000-2999
2001:0:5::1	2.3.4.5	1500-3999

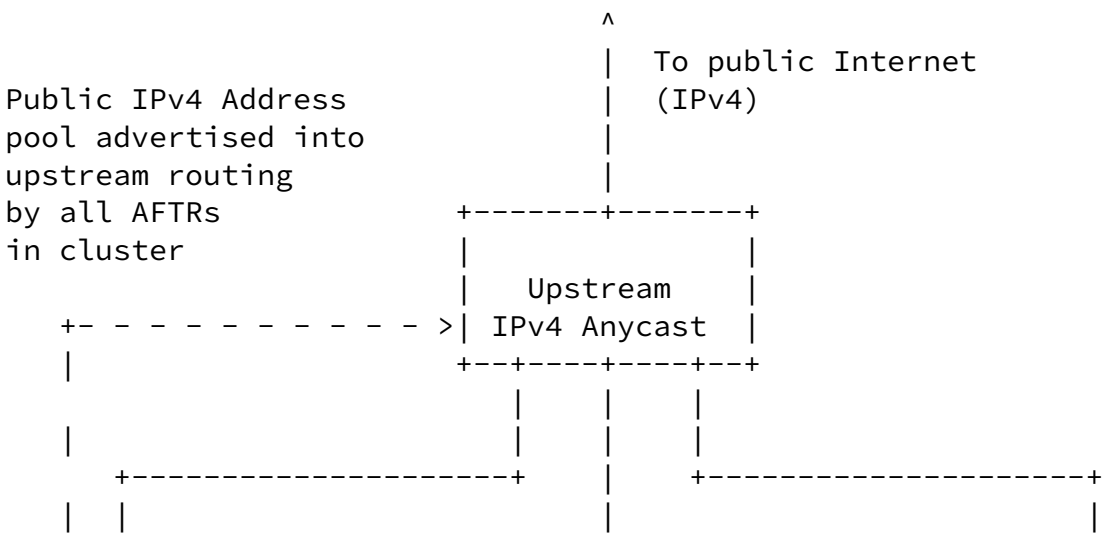
Figure 1 DHCP4o6/AFTR Per-subscriber configuration table data example

This memo proposes a simple architecture to guarantee the synchronization of those mapping tables and rely on anycast IPv4 and IPv6 technologies to provide failover within the AFTRs in the same

cluster.

2.4. IPv6 and IPv4 Anycast Considerations

The following diagram shows the architecture for the Lightweight 4over6 cluster deployment.



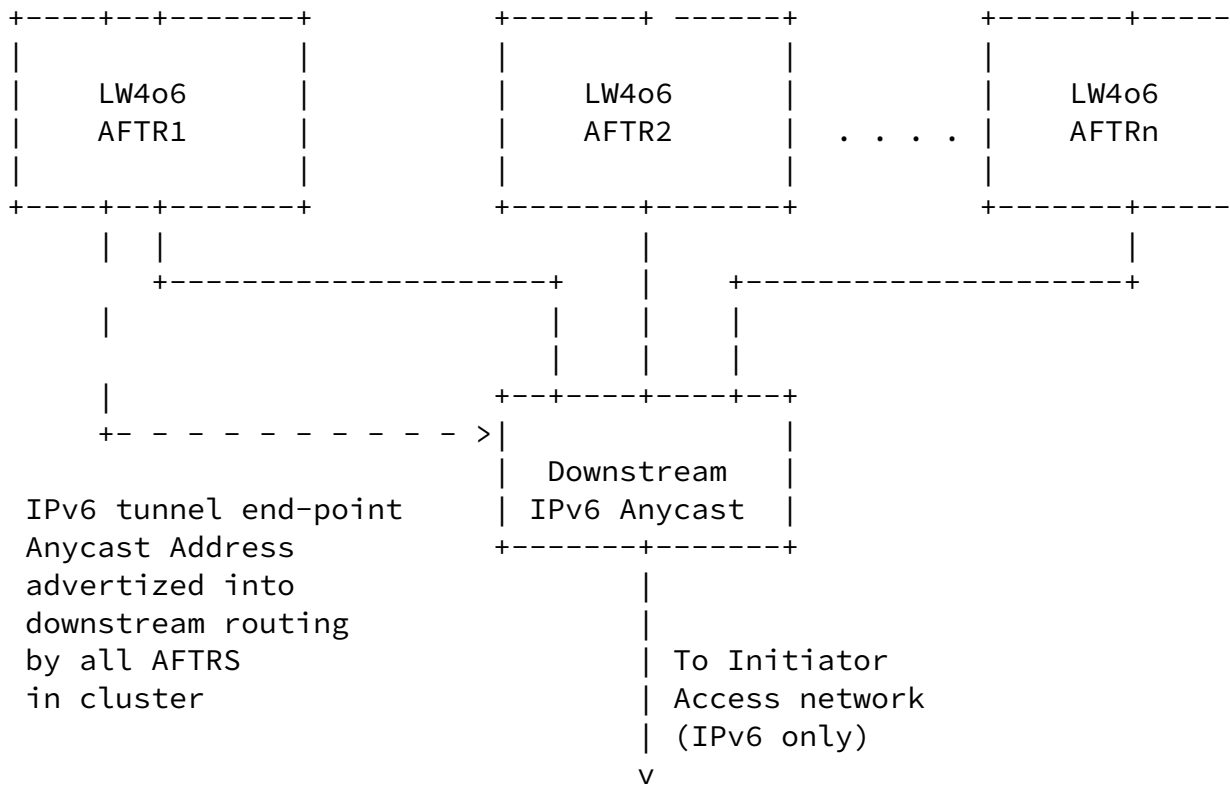


Figure 2 Lightweight 4over6 Cluster

To increase service availability, A simple way to achieve fail-over is to configure both the IPv6 tunnel end point address and the IPv4

address pool as anycast addresses on the AFTRs and announce these routes into the IGP run by the ISP, such as OSPF or IS-IS.

The number of AFTRs in a cluster provides the degree of redundancy of the solution. In practice, two AFTRs are expected to be sufficient in most cases.

2.5. Load Balancing across Multiple Concentrators

AFTR functionality can be scaled by load balancing the encapsulation/ decapsulation across multiple AFTRs in a cluster. Due to the commonality of configuration and stateless nature of the solution, any tunneled packet from any Initiator served by the cluster can arrive at any cluster member and will be processed in the same way. Likewise, for inbound packets originating in the IPv4 realm, a packet

that arrives at any of the cluster member will be encapsulated and sent to the correct initiator.

Load balancing could be achieved using specific load balancing infrastructure to distribute the tunnels and inbound v4 traffic across the cluster. It is also possible to use the Equal Cost Multipath inherent in some routing protocols to achieve this.

In order to prevent out-of-sequence packets in the tunnelled traffic, a mechanism for forwarding all packets belonging to a single tunnel through the same cluster member should be used. An example of this would be a source/destination hashing algorithm such as [RFC2992] describes.

2.6. DHCPv6 Tunnel End-point Option Considerations

All CPEs belonging to the same group of subscribers need to receive the same tunnel end-point option (via DHCPv6). This will be set to the IPv6 anycast address of the AFTR cluster.

2.7. CPE IPv6 Address Management

The DHCPv4 server uses the IPv6 address of the CPE as its index. In order to keep the overall service architecture flexible and adaptable, it is preferable that the CPE is configured using DHCPv6 out of a specific pool reserved by the ISP.

2.8. Binding Table Synchronization

It is proposed that binding tables be pre-computed and stored statically on the AFTRs and the DHCPv4 servers. The method of creating the binding tables is out of the scope of this memo.

These tables are not expected to change regularly. Typical reasons for an update include adding or removing an IPv4 address block, or changing the size of IPv4 ports ranges available to each CPE.

To ensure continuous operation, binding tables have to be updated simultaneously across all AFTRs in a cluster by a mechanism such as netconf. It may also be necessary to reconfigure CPEs during this process (e.g. via a DHCPv6 reconifgure message). The details are out

of scope for this memo.

2.9. Subscriber Management and Growth

It is recommended that the ISP predefines all IPv6 addresses and corresponding IPv4 addresses and port ranges for any given subscriber group.

If the ISP runs out of space within a subscriber group, another group is then defined. Cutomer CPEs can be migrated between different subscriber groups by alterning the CPE configuration over DHCP.

2.10. Privacy Extensions

In some deployments, regulations require that IP addresses allocated to customers can be changed periodically or on demand to protect users privacy.

This can be achieve by rolling over the IPv6 addresses in the DHCPv6 server allocating IPv6 addresses to the CPE. If all subscribers within the subscriber group are allocated the same number of ports in IPv4, then the IPv6 to IPv4 address and port binding may remain the same, the IPv4 address and ports will then roll over automatically at the same time as the IPv6 addresses do.

[I-D.cui-softwire-b4-translated-ds-lite] states that when the IPv6 address of the B4 is changed, then DHCPv4over6 configuration process must be re-initiated.

3. IANA Considerations

None.

4. Security Considerations

None.

5. Acknowledgements

6. References

6.1. Normative References

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

[RFC2992] Hopps, C., "Analysis of an Equal-Cost Multi-Path Algorithm", RFC 2992, November 2000.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.

6.2. Informative References

[I-D.cui-softwire-b4-translated-ds-lite] Cui, Y., Sun, Q., Boucadair, M., Tsou, T., Lee, Y., and I. Farrer, "Lightweight 4over6: An Extension to the DS-Lite Architecture", draft-cui-softwire-b4-translated-ds-lite-07 (work in progress), July 2012.

Authors' Addresses

Ian Farrer
Deutsche Telekom AG
GTN-FM4
Landgrabenweg 151
Bonn 53227
Germany

Email: ian.farrer@telekom.de

Alain Durand
Juniper Networks
1194 North Mathilda Avenue
Sunnyvale, CA 94089-1206
USA

Email: adurand@juniper.net