Softwire
Internet-Draft
Intended status: Informational
Expires: April 18, 2011

Y. Lee
Comcast
R. Maglione
Telecom Italia
C. Williams
MCSR Labs
C. Jacquenet
M. Boucadair
France Telecom

October 15, 2010

Deployment Considerations for Dual-Stack Lite
draft-lee-softwire-dslite-deployment-00

Abstract

This document discusses the deployment issues and describes requirements for the deployment and operation of Dual-Stack Lite. This document describes the various deployment scenarios and applicability of the Dual-Stack Lite protocol.

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 18, 2011.

Copyright Notice

Table of Contents

1. Overview

Dual-stack Lite (DS-Lite) [I-D.ietf-softwire-dual-stack-lite] is a transition technique that enable operators to multiplex public IPv4 addresses while provisioning only IPv6 to users. DS-Lite is designed to address the IPv4 depletion issue and allow the operators to upgrade their network incrementally to IPv6. DS-Lite combines IPv4-in-IPv6 tunnel and NAT44 to share a public IPv4 address more than one user. This document discusses various deployment considerations for DS-Lite by operators.

2. AFTR Deployment Considerations

Address Family Transition Router (AFTR) is the function deployed inside the operator's network. AFTR can be a standalone device or embedded into a router. AFR is the IPv4-in-IPv6 tunnel termination point and the NAT44 device. It is deployed at the IPv4-IPv6 network border where the tunnel interface is IPv6 and the NAT interface is IPv4. Although an operator can configure a dual-stack interface for both functions, we strongly recommended to configure two individual interfaces (i.e. one dedicated for IPv4 and one dedicated for IPv6) to segregate the functions.

In this section, the deployment considerations for AFTR are described.

2.1. MTU Considerations

DS-Lite is part tunneling protocol. Tunneling introduces some additional complexity and has a risk of MTU or other mis-configurations. With tunneling comes additional header overhead that implies that the tunnel's MTU is smaller than the raw interface MTU. The second problem is that between the B4 and AFTR networking entities there may exist further tunnels inside tunnel, so that the tunnel ingress is not necessarily aware of the true tunnel MTU. The third problem is that the routing of the interior of the tunnel may change, so that the tunnel MTU may be variable. The issue that the end user will experience is that they cannot download Internet pages or transfer files using File Transfer Protocol (FTP) but may be able to ping successfully.

For fragmentation problem shares among all the tunneling protocols, this is not unique to DS-Lite. The IPv4 packet isn't over-sized, it is the v6 encapsulation that MAY cause the oversized issue. So the tunnel points are responsible to handle the fragmentation. In general, the Tunnel-Entry Point and Tunnel-Exist Point should fragment and reassemble the oversize datagram. This mechanism is

transport protocol agnostic and work for both UDP and TCP. For TCP, we could potentially avoid fragmentation by modify MSS option. The B4 networking component may send an ICMP Destination Unreachable-Fragmentation Needed and DF Set message back to the sending host in the subscriber network.

2.2. Lawful Intercept Considerations

Because of its IPv4-in-IPv6 tunneling scheme, interception in DS-Lite architecture must be performed on the AFTR itself. Timestamped logging of the address and port mappings at the AFTR must be maintained, which in turn can add a heavy resource burden to the AFTR devices.

Logging to a storage device off the AFTR may also contribute to network load. Wiretapping of a single subject may mean statically mapping the user to a certain range of ports on a single address, to remove the need to follow dynamic port mappings. A single IPv4 address, or some range of ports for each address, might be set aside for wiretapping purposes to simplify such procedures. But any requirement that users behind a given AFTR be logged is going to mean logging not only traffic but all changes to the mapping tables.

2.3. Logging at the AFTR

The timestamped logging of address and port mappings is essential not only for lawful intercept but also for tracing back specific users when a problem is identified from the outside of the AFTR. Such a problem is usually a misbehaving user in the case of a spammer or a DoS source, or someone violating a usage policy. Knowing the user may result in black-listing. Without time-specific logs of the address and port mappings, a misbehaving user stays well hidden behind the AFTR.

Blacklisting might restrict others in the home or office from accessing the website but altogether few innocent bystanders are affected. What happens, though, if a website bans an IPv4 address on the outside of an AFTR? In the effort to restrict a single user, hundreds of people may be inadvertently restricted generally all subscribers on a CMTS or a group of BNASes behind the AFTR.

Black- or white-listing may need to be split in an AFTR architecture. Polices applying to incoming sources must be implemented on the outside of the AFTR. Once the packets are translated, they cannot be easily identified by IPv4 address without some correlation with the AFTR mapping table.

2.3.1. AFTR's Policies

Polices applying on the NAT-ed addresses must be implemented on the external interface of the AFTR. Once the packets are translated, they cannot be easily identified by IPv4 address without some correlation with the AFTR mapping table. Policies applying to outgoing sources must be implemented on the customer-facing side of the AFTR for the same reason. In order to be able to deploy different services offers, multiple set of policies (e.g. QoS and ACL settings) can be configured on the AFTR: each set of policies can then be applied to a different logical tunnel interface on the ATFR.

2.4. AFTR Impacts on Internal Accounting Systems

Single points of failure, potential address pool depletion attacks, performance and scalability, effects on fragmented packets, effects on asymmetric traffic flows, required modifications to provisioning systems, required modifications to internal accounting systems.

2.4.1. AFTR Impacts on Accounting Process in Broadband Access

DS-Lite introduces challenges to IPv4 accounting process. In a typical DSL/Broadband access scenario where the Residential Gateway (RG) is acting as a B4 element, the BNAS is the IPv6 edge router which connects to the AFTR. The BNAS is normally responsible for IPv6 accounting and all the subscriber manager functions such as authentication, authorization and accounting. However, given the fact that IPv4 traffic is encapsulated into an IPv6 packet at the B4 level and only decapsulated at the ATFR level, the BNAS can't do the IPv4 accounting without examining the inner packet. AFTR is the next logical place to perform IPv4 accounting, but it will potentially introduce some additional complexity because the AFTR does not have detailed customer identity information.

The accounting process at the AFTR level is only necessary if the Service Provider requires separate per user accounting records for IPv4 and IPv6 traffic. If the per user IPv6 accounting records, collected by the BNAS, are sufficient, the additional complexity to be able to implement IPv4 accounting at the ATFR level is not required. It is important to consider that, since the IPv4 traffic is encapsulated in IPv6 packets, the data collected by the BNAS for IPv6 traffic already contain the total amount of traffic (i.e. IPv6 plus IPv4).

Even if detailed accounting records collection for IPv4 traffic may not be required, in some scenarios it would be useful for a Service Provider, to have inside the RADIUS Accounting packet, generated by the BNAS for the IPv6 traffic, a piece of information that can be

used to identify the AFTR that is handling the IPv4 traffic for that user. This can be achieved by adding into the IPv6 accounting records the RADIUS attribute information specified in [I-D.ietf-softwire-dslite-radius-ext]

2.5. Reliability Considerations of AFTR

The service provider can use techniques to achieve high availability such as various types of clusters to ensure availability of the IPv4 service. High availability techniques include the cold standby mode. In this mode the AFTR states are not replicated from the Primary AFTR to the Backup AFTR. When the Primary AFTR fails, all the existing established sessions will be flushed out. The internal hosts are required to re-establish sessions to the external hosts. Another high availability option is the hot standby mode. In this mode the AFTR keeps established sessions while failover happens. AFTR states are replicated from the Primary AFTR to the Backup AFTR. When the Primary AFTR fails, the Backup AFTR will take over all the existing established sessions. In this mode the internal hosts are not required to re-establish sessions to the external hosts. The final option is to deploy a mode in between these two whereby only selected sessions such as critical protocols are replicated. Criteria for sessions to be replicated on the backup would be explicitly configured on the AFTR devices of a redundancy group.

2.6. Strategic Placement of AFTR

The public IPv4 addresses are pulled away from the customer edge to the outside of the centralized AFTR where many customer networks can share a single public IPv4 address.

The AFTR architecture design, then, is mostly figuring out the strategic placement of each AFTR to best use the capacity of each public IPv4 address without oversubscribing the address or overtaxing the AFTR itself. Although only a few studies of per-user port usage have been done, an AFTR should be able to support 3000 - 5000 users per public IPv4 address.

By centralizing public IPv4 addresses, each address no longer represents a single machine, a single household, or a single small office. The address now represents thousands of machines, homes, and offices related only in that they are behind the same AFTR. Identification by IP address becomes difficult or impossible and thus applications that assume such geographic information may not work as intended.

2.7. AFTR Considerations for Geographically Aware Services

Various applications and services will place their servers in such a way to locate them near sets of user so that this will lessen the latency on the client end. In addition, having sufficient geographical coverage can indirectly improve end-to-end latency. An example is that nameservers typically return results optimized for the DNS resolver's location. Deployment of AFTR must be done in such a way as not to negatively impact the geographical nature of these services. This can be done by making sure that AFTR deployments are geographically distributed so that existing assumptions of the clients source IP address by geographically aware servers can be maintained.

2.8. Impacts on QoS

As with tunneling in general there are challenges with deep packet inspection with DS-Lite for purposes of QoS. Service Providers commonly uses DSCP to classify and prioritize packets. It is recommended the AFTR to copy the DSCP value in the IPv4 header to the IPv6 header after the encapsulation.

2.9. Port Forwarding Considerations

Some applications require accepting incoming UDP or TCP traffic. When the remote host is on IPv4, the incoming traffic will be directed towards an IPv4 address. Some applications use (UPnP-IGD) (e.g., XBox) or ICE [I-D.ietf-mmusic-ice] (e.g., SIP, Yahoo!, Google, Microsoft chat networks), other applications have all but completely abandoned incoming connections (e.g., most FTP transfers use passive mode). But some applications rely on ALGs, UPnP IGD, or manual port configuration. Port Control Protocol (PCP) [I-D.wing-pcp-design-considerations] is designed to address this issues.

3. B4 Deployment Considerations

In order to configure the IPv4-in-IPv6 tunnel, the B4 element needs the IPv6 address of the AFTR element. This IPv6 address can be configured using a variety of methods, ranging from an out-of-band mechanism, manual configuration or a variety of DHCPv6 options. In order to guarantee interoperability, a B4 element SHOULD implement the DHCPv6 option defined in [I-D.ietf-softwire-ds-lite-tunnel-option]. The DHCP server must be reachable via normal DHCP request channels from the B4, and it must be configured with the AFTR address. In Broadband Access scenario where AAA/RADUIS is used for provisioning user profiles in the BNAS,

[I-D.ietf-softwire-dslite-radius-ext] may be used. BNAS will learn the AFTR address from the RADIUS attribute and act as the DHCPv6 server for the B4s.

3.1. DNS deployment Considerations

[I-D.ietf-softwire-dual-stack-lite] recommends configuring the B4 with a DNS proxy resolver, which will forward queries to an external recursive resolver over IPv6. Alternately, the B4 proxy resolver can be statically configured with the IPv4 address of an external recursive resolver. In this case, DNS traffic to the external resolver will be tunneled through IPv6 to the AFTR. Note that the B4 must also be statically configured with an IPv4 address in order to source packets; the draft recommends an address in the 192.0.0.0/29 range. Even more simply, you could eliminate the DNS proxy, and configure the DHCP server on the B4 to give its clients the IPv4 address of an external recursive resolver. Because of the extra traffic through the AFTR, and because of the need to statically configure the B4, these alternate solutions are likely to be unsatisfactory in a production environment. However, they may be desirable in a testing or demonstration environment.

4. Security Considerations

This document does not present any new security issues. [I-D.ietf-softwire-dual-stack-lite] discusses DS-Lite related security issues. General NAT security issues are not repeated here.

Some of the security issues with carrier-grade NAT result directly from the sharing of the routable IPv4 address. Addresses and timestamps are often used to identify a particular user, but with shared addresses, more information (i.e., protocol and port numbers) is needed. This impacts software used for logging and tracing spam, denial of service attacks, and other abuses. Devices on the customers side may try to carry out general attacks against systems on the global Internet or against other customers by using inappropriate IPv4 source addresses inside tunneled traffic. The AFTR needs to protect against such abuse. One customer may try to carry out a denial of service attack against other customers by monopolizing the available port numbers. The AFTR needs to ensure equitable access. At a more sophisticated level, a customer may try to attack specific ports used by other customers. This may be more difficult to detect and to mitigate without a complete system for authentication by port umber, which would represent a huge security requirement.

5. Conclusion

DS-Lite provides new functionality to transition IPv4 traffic to IPv6 addresses. As the supply of unique IPv4 addresses diminishes, service providers can now allocate new subscriber homes IPv6 addresses and IPv6-capable equipment. DS-Lite provides a means for the private IPv4 addresses behind the IPv6 equipment to reach the IPv4 network.

This document discusses the issues that arise when deploying DS-Lite in various deployment modes. Hence, this document can be a useful reference for service providers and network designers. Deployment considerations of the B4, AFTR and DNS have been discussed and recommendations for their usage have been documented.

6. Acknowledgement

TBD

7. IANA Considerations

This memo includes no request to IANA.

8. References

8.1. Normative References

[I-D.ietf-softwire-ds-lite-tunnel-option]
Hankins, D. and T. Mrugalski, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6) Option for Dual- Stack Lite", draft-ietf-softwire-ds-lite-tunnel-option-05 (work in progress), September 2010.

[I-D.ietf-softwire-dslite-radius-ext]
Maglione, R. and A. Durand, "RADIUS Extensions for Dual-Stack Lite", draft-ietf-softwire-dslite-radius-ext-00 (work in progress), October 2010.

[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.

[I-D.wing-pcp-design-considerations]

Wing, D., "PCP Design Considerations", draft-wing-pcp-design-considerations-00 (work in progress), September 2010.

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

[RFC4925] Li, X., Dawkins, S., Ward, D., and A. Durand, "Softwire Problem Statement", RFC 4925, July 2007.

8.2. Informative References

[I-D.ietf-mmusic-ice]
Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", draft-ietf-mmusic-ice-19 (work in progress), October 2007.

[I-D.ietf-v6ops-ipv6-cpe-router]
Singh, H., Beebee, W., Donley, C., Stark, B., and O. Troan, "Basic Requirements for IPv6 Customer Edge Routers", draft-ietf-v6ops-ipv6-cpe-router-07 (work in progress), August 2010.

[I-D.xu-behave-stateful-nat-standby]
Xu, X., "Redundancy and Load Balancing Framework for Stateful Network Address Translators (NAT)", draft-xu-behave-stateful-nat-standby-05 (work in progress), September 2010.

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

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

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

[RFC5569] Despres, R., "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd)", RFC 5569, January 2010.

Authors' Addresses

Yiu L. Lee
Comcast
One Comcast Center
Philadelphia, PA 19103
U.S.A.

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

Roberta Maglione
Telecom Italia
Via Reiss Romoli 274
Torino 10148
Italy

Email: roberta.maglione@telecomitalia.it
URI: http://www.telecomitalia.it

Carl Williams
MCSR Labs
Philadelphia
U.S.A.

Email: carlw@mcsr-labs.org

Christian Jacquenet
France Telecom
Rennes
France

Email: christian.jacquenet@orange-ftgroup.com>

Mohamed Boucadair
France Telecom
Rennes
France

Email: mohamed.boucadair@orange-ftgroup.com