Softwire
Internet-Draft
Intended status: Informational
Expires: January 12, 2012

Y. Lee
Comcast
R. Maglione
Telecom Italia
C. Williams
MCSR Labs
C. Jacquenet
M. Boucadair
France Telecom
July 11, 2011

Deployment Considerations for Dual-Stack Lite
draft-lee-softwire-dslite-deployment-02

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 considerations and applicability of the Dual-Stack Lite architecture.

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 12, 2012.

Copyright Notice

Table of Contents

1. Overview 3
2. AFTR Deployment Considerations 3
2.1. Interface Consideration 3
2.2. MTU Considerations 3
2.3. Fragmentation 3
2.4. Lawful Intercept Considerations 4
2.5. Logging at the AFTR 4
2.6. Blacklisting a shared IPv4 Address 5
2.7. AFTR's Policies 5
2.8. AFTR Impacts on Accounting Process in Broadband Access . . 5
2.9. Reliability Considerations of AFTR 6
2.10. Strategic Placement of AFTR 6
2.11. AFTR Considerations for Geographically Aware Services . . 7
2.12. Impacts on QoS 8
2.13. Port Forwarding Considerations 8
2.14. DS-Lite Tunnel Security 8
2.15. IPv6-only Network considerations 9
3. B4 Deployment Considerations 9
3.1. DNS deployment Considerations 10
4. Security Considerations 10
5. Conclusion 10
6. Acknowledgement 11
7. IANA Considerations 11
8. References 11
8.1. Normative References 11
8.2. Informative References 12
Authors' Addresses 13

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

2.1. Interface Consideration

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. AFTR 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 recommended to configure two individual interfaces (i.e. one dedicated for IPv4 and one dedicated for IPv6) to segregate the functions.

2.2. MTU Considerations

DS-Lite is part tunneling protocol. Tunneling introduces some additional complexity and has a risk of MTU. With tunneling comes additional header overhead that implies that the tunnel's MTU is smaller than the raw interface MTU. The issue that the end user may experience is that they cannot download Internet pages or transfer files using File Transfer Protocol (FTP).

To mitigate the tunnel overhead, the access network could increase the MTU size to account the necessary tunnel overhead which is the size of an IPv6 header. If the access network MTU size is fixed and cannot be changed, the B4 element and the AFTR must support fragmentation.

2.3. Fragmentation

The IPv4-in-IPv6 tunnel is between B4 and AFTR. When a host behind the B4 element communicates to a server, both the host and the server are not aware of the tunnel. They may continue to use the maximum MTU size for communication. In fact, the IPv4 packet isn't over-sized, it is the v6 encapsulation that may cause the oversize. 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 oversized datagram. If the DF is set, the B4 element should send an ICMP "Destination Unreachable" with "Fragmentation Needed and Don't Fragment was Set" and drop the packet. If the DF is not set, the B4 element should fragment the IPv6 packet after the encapsulation. This mechanism is transport protocol agnostic and works for both UDP and TCP.

[editor note: Should we drop the IPv4 packet when DF is set?]

2.4. Lawful Intercept Considerations

Because of its IPv4-in-IPv6 tunneling scheme, interception of IPv4 sessions in DS-Lite architecture must be performed on the AFTR. Subjects can be uniquely identified by the IPv6 address assigned to the B4 element. Operator must associate the B4's IPv6 address and the public IPv4 address and port used by the subject.

Monitoring of a single subject may mean statically mapping the subject to a certain range of ports on a single IPv4 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 monitoring purposes to simplify such procedures. This requires to create a static mapping of a B4 element's IPv6 address to an IPv4 address that used for lawful intercept.

2.5. Logging at the AFTR

The timestamped logging is essential 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. Without time-specific logs of the address and port mappings, a misbehaving user stays well hidden behind the AFTR.

In DS-Lite framework, each B4 element is given a unique IPv6 address. The AFTR uses this IPv6 address to identify the B4 element. Thus, the AFTR must log the B4's IPv6 address and the IPv4 information. There are two types of logging: (1) Source-Specific Log and (2) Destination-Specific Log. For Source-Specific Log, the AFTR must timestamped log the B4's IPv6 address, transport protocol, source IPv4 address after NAT-ed, and source port. If a range of ports is dynamically assigned to a B4 element, the AFTR may create one log per range of ports to aggregate number of log entries. For Destination-Specific Log, the AFTR must timestamped log the B4's IPv6 address, transport protocol, source IPv4 address after NAT-ed, source port, destination address and destination port. The AFTR must log every

session from the B4 elements. No log aggregation can be performed. When using Destination-Specific Log, the operator must be careful of the large number of log entries created by the AFTR.

2.6. Blacklisting a shared IPv4 Address

AFTR is a NAT device. It shares a single IPv4 address with multiple users. [I-D.ietf-intarea-shared-addressing-issues] discusses many considerations when sharing address. When a pubic IPv4 address is blacklisted, this may affect multiple users and there is no effective way for the B4 element to notify the AFTR an IP address is blacklisted. It is recommended the server must no longer rely solely on IP address to identify an abused user. The server should combine the information stored in the transport layer (e.g. source port) and application layer (e.g. HTTP) to identify an abused user. [I-D.boucadair-intarea-nat-reveal-analysis] analyzes different approaches to identify a user in a shared address environment.

2.7. AFTR's Policies

There are two types of AFTR polices: (1) Outgoing Policies and (2) Incoming Policies. The outgoing policies must be implemented on the AFTR's internal interface connected to the B4 elements. The policies may include ACL and QoS settings. For example: the AFTR may only accept B4's connections originated from the IPv6 prefixes provisioned in the AFTR. The AFTR may also give priority to the packets marked by certain DSCP values. The AFTR may also limit the rate of port creation from a single B4's IPv6 address. Outgoing policies could be applied to individual B4 element or a set of B4 elements.

The incoming policies must be implemented on the AFTR's external interface connected to the IPv4 network. Similar to the outgoing policies, the policies may include ACL and QoS settings. Incoming policies are usually more general and globally applied to all users rather than individual user.

2.8. 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]

2.9. 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.

2.10. 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 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.

AFTR is a tunnel concentrator, B4 traffic must pass through the AFTR to reach the IPv4 Internet. Managing tunnels and NAT could be resource intensive, so the placement of the AFTR would affect the traffic flows in the access network and have operation implications. In general, there are two placements to deploy AFTR. Model One is to deploy the AFTR in the edge of network to cover a small region. Model Two is to deploy the AFTR in the core of network to cover a large region.

When the operator consider where to deploy the AFTR, they must make trade-offs. AFTR in Model One serves few B4 elements, thus, it requires less powerful AFTR. Moreover, the traffic flows are more evenly distributed to the AFTRs. However, it requires to deploy more AFTRs to cover the entire network. Often the operation cost increases proportionally to the number of network equipment. AFTR in Model Two covers larger area, thus, it serves more B4 elements. The operator could deploy only few AFTRs in the strategic locations to support the entire subscriber base. However, this model requires more powerful AFTR to sustain the load at peak hours. Since the AFTR would support B4 elements from different regions, the AFTR would be deployed deeper in the network and steer more traffic flows to the network where the AFTR is located.

DS-Lite framework can be incrementally deployed. An operator may consider to start with Model Two. When the demand increases, they could push the AFTR closer to the edge which would effectively become Model One.

2.11. AFTR Considerations for Geographically Aware Services

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 hundreds of machines, homes, and offices related only in that they are behind the same AFTR. Identification by IP address becomes more difficult and thus applications that assume such geographic information may not work as intended.

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 could 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. Another possibility the application could rely on location information such as GPS co-ordination to identify the user's location. This technique is commonly used in mobile deployment where the mobile devices are probably behind a NAT device.

2.12. 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 different types of traffic. DS-Lite tunnel can be seen as particular case of uniform conceptual tunnel model described in section 3.1 of [RFC2983]. The uniform model views an IP tunnel as just a necessary mechanism to get traffic to its destination, but the tunnel has no significant impact on traffic conditioning. In this model, any packet has exactly one DS Field that is used for traffic conditioning at any point and it is the field in the outermost IP header. In DS-Lite model this is the Traffic Class field in IPv6 header. According to [RFC2983] implementations of this model copy the DS value to the outer IP header at encapsulation and copy the outer header's DSCP value to the inner IP header at decapsulation. Applying the described model to DS-Lite scenario, it is recommended that the AFTR propagates the DSCP value in the IPv4 header to the IPv6 header after the encapsulation for the downstream traffic and, in the same way, the B4 propagates the DSCP value in the IPv4 header to the IPv6 header after the encapsulation for the upstream traffic.

2.13. 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.ietf-pcp-base] is designed to address this issues.

2.14. DS-Lite Tunnel Security

Section 11 of [I-D.ietf-softwire-dual-stack-lite] describes security issues associated to DS-Lite mechanism. One of the recommendations contained in this section, in order to limit service offered by AFTR only to registered customers, is to implement IPv6 ingress filter on

the AFTR's tunnel interface to accept only the IPv6 address range defined in the filter. This approach requires to know in advance the IPv6 prefix delegated to the customers in order to be able to configure the filter.

An alternative way to achieve the same goal and to provide some form of access control to the DS-Lite tunnel, is to use DHCPv6 Leasequery defined in [RFC5007]. When the AFTR receives a packet from an unknown (new) prefix it issues a DHCPv6 Leasequery based on IPv6 address to the DHCPv6 server in order to verify if that prefix was previously delegated by the DHCPv6 server to that specific client. The DHCPv6 Server will reply with the delegated prefix and the associated lease. If the two prefix are the same the ATFR accepts the packet otherwise it drops it and it denies the service.

2.15. IPv6-only Network considerations

In environments where the service provider wants to deploy AFTR in the IPv6 core network, the AFTR nodes may not have direct IPv4 connectivity. In this scenario the service provider extends the IPv6-only boundary to the border of the network and only the border routers have IPv4 connectivity. For both scalability and performance purposes AFTR capabilities are located in the IPv6-only core closer to B4 elements. The service provider assigns only IPv6 prefixes to the B4 capable devices but also continues to provide IPv4 services to these customers. In this scenario the AFTR has only IPv6-connectivity and must be able to send and receive IPv4 packets. Enhancements to the DS-LITE AFTR are required to achive this. [I-D.boucadair-softwire-dslite-v6only] describes such issues and enhancements to DS-Lite in IPv6-only deployments.

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-pcp-base]
Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", draft-ietf-pcp-base-13 (work in progress), July 2011.

[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-10 (work in progress), March 2011.

[I-D.ietf-softwire-dslite-radius-ext]
Maglione, R. and A. Durand, "RADIUS Extensions for Dual-Stack Lite", draft-ietf-softwire-dslite-radius-ext-02 (work in progress), March 2011.

[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-11 (work in progress), May 2011.

[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.

[RFC5007] Brzozowski, J., Kinnear, K., Volz, B., and S. Zeng, "DHCPv6 Leasequery", RFC 5007, September 2007.

8.2. Informative References

[I-D.boucadair-intarea-nat-reveal-analysis]
Boucadair, M., Touch, J., Levis, P., and R. Penno, "Analysis of Solution Candidates to Reveal a Host Identifier in Shared Address Deployments", draft-boucadair-intarea-nat-reveal-analysis-03 (work in progress), June 2011.

[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-intarea-server-logging-recommendations]
Durand, A., Gashinsky, I., Lee, D., and S. Sheppard, "Logging recommendations for Internet facing servers", draft-ietf-intarea-server-logging-recommendations-04 (work in progress), April 2011.

[I-D.ietf-intarea-shared-addressing-issues]
Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", draft-ietf-intarea-shared-addressing-issues-05 (work in progress), March 2011.

[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-09 (work in progress), December 2010.

[I-D.xu-behave-stateful-nat-standby]
Xu, X., Boucadair, M., Lee, Y., and G. Chen, "Redundancy Requirements and Framework for Stateful Network Address Translators (NAT)", draft-xu-behave-stateful-nat-standby-06 (work in progress), October 2010.

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

[RFC2983] Black, D., "Differentiated Services and Tunnels", RFC 2983, October 2000.

[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:

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