Softwire   Y. Lee
Internet-Draft   Comcast
Intended status: Informational   R. Maglione
Expires: September 5, 2012   Telecom Italia
                                                        C. Williams
                                                          MCSR Labs
                                                       C. Jacquenet
                                                       M. Boucadair
                                                     France Telecom
                                                      March 4, 2012

               Deployment Considerations for Dual-Stack Lite
                  draft-ietf-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 September 5, 2012.

Copyright Notice

Table of Contents

1. Overview

   Dual-stack Lite (DS-Lite) [RFC6333] is a transition technique that
   enable operators to multiplex public IPv4 addresses while
   provisioning only IPv6 to users. DS-Lite is designed to continue
   offering IPv4 services while operators upgrading their network
   incrementally to IPv6. DS-Lite combines IPv4-in-IPv6 [RFC2473]
   softwire and NAT44 [RFC1631] to enable more than one user to share a
   public IPv4 address. This document discusses various DS-Lite
   deployment considerations for operators.

2. AFTR Deployment Considerations

2.1. Interface Consideration

   Address Family Transition Router (AFTR) is a network element that
   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
   external NAT44 interface is IPv4. Although an operator can configure
   a dual-stack interface for both functions, we recommend 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 overhead to
   the packet and decreases the effective MTU size after encapsulation.
   The DS-lite users may experience problems with applications such as
   not being able to download Internet pages or transfer large file. To
   mitigate the tunnel overhead, the access network may increase the MTU
   size to account the necessary tunnel overhead. If simple IPv4-in-
   IPv6 softwire [RFC2473] is used, the overhead 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
   defined in [RFC6333].

2.3. Fragmentation

   The IPv4-in-IPv6 tunnel is established between B4 and AFTR. When a
   host behind the B4 element communicates with a remote peer, both end
   nodes are not aware of the tunnel. For example, the peers may use
   the MTU size associated with their connected interfaces. In fact,
   the IPv4 packet isn't over-sized, it is the IPv6 encapsulation that
   may cause the oversize of the encapsulating packets. So the tunnel
   endpoints are responsible for handling the fragmentation. In

general, the Tunnel-Entry Point and Tunnel-Exit Point should fragment
and reassemble the oversized datagram. If the DF bit is set in the
IPv4 header, 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 unset in the IPv4 header, the B4
element should fragment the IPv6 packet after the encapsulation.
This mechanism is transport protocol agnostic and works for transport
protocol such as TCP and UDP over IP.

2.4. Lawful Intercept Considerations

Because of the IPv4-in-IPv6 tunneling scheme, interception of IPv4
sessions in DS-Lite framework is likely performed on the AFTR.
Subjects can be uniquely identified by the IPv6 address assigned to
the B4 element. If an operator is legally requested to intercept
packets of a subject, the AFTR should extract the IPv4 packets from
the IPv6 payload before sending it to the interception point.

Monitoring of a subject may require statically mapping the subject to
a certain range of ports of 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 creating a
static mapping of a B4 element's IPv6 address to a public IPv4
address and port range that are used for lawful intercept.

2.5. Logging at the AFTR

Timestamped logging is essential for back tracking specific users
when a problem is identified with one of the AFTR's NAT-ed addresses.
Such a problem is usually a misbehaving user in the case of a spammer
or a Deny-of-Service (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 provisioned with one or more
unique source IPv6 addresses. The AFTR uses the B4's tunnel IPv6
address to identify the B4 element. Thus, to uniquely identify a
specific user, the AFTR is required to log more than just IPv4
address. There are two types of logging:

o Source-Specific Log

o 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
create a timestamped log of the B4's IPv6 address, transport
protocol, source IPv4 address after NAT-ed, source port, destination
address and destination port.

Destination-Specific Log is session-based, the operators can't really
aggregate log entries. When using Destination-Specific Log, the
operator must be careful of the large number of log entries created
by the AFTR. Destination-Specific Log may raise privacy concerns.
Operators should apply the same privacy policies for both regular and
DS-Lite users.

2.6. Blacklisting a shared IPv4 Address

AFTR is a NAT device. It enables multiple users to share a single
public IPv4 address. [RFC6269] discusses some considerations when
sharing an IPv4 address. When a public IPv4 address is blacklisted
by a remote peer, this may affect multiple users. Internet hosts
such as servers 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 [RFC6302].
[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:

o Outgoing Policies

o Incoming Policies

The outgoing policies should be implemented on the AFTR's internal
interface connected to the B4 elements. The policies may include
Access Control List (ACL) and Qualify of Service (QoS) settings. For
example: the AFTR may only accept B4's connections originated from
the IPv6 prefixes configured in the AFTR. The AFTR may also give
priority to the packets marked by certain DSCP values [RFC2475]; the
AFTR may also limit the rate of port allocation for a single B4's
IPv6 address.

An operator may create multiple outgoing polices in the AFTR, each
identified by a softwire. When provisioning a user, the system will
pass the softwire identifier associated to a specific incoming policy
to the user. Two standardized mechanisms to pass softwire identifier
to the B4 element are DHCPv6 [RFC6333] and RADIUS [RFC6519].

Outgoing policies could be applied to an individual B4 element or to a set of B4 elements.

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

2.8. AFTR Impacts on Accounting Process

DS-Lite introduces challenges to IPv4 accounting process. In a typical broadband access scenario (e.g. DSL or Cable), the B4 element is embedded in the Residential Gateway and the edge router (e.g. BRAS or CMTS) is the IPv6 edge router. The edge router is usually responsible for IPv6 accounting and the subscriber management functions such as authentication, authorization and accounting. However, given the fact that IPv4 traffic is encapsulated in an IPv6 packet at the B4 and only decapsulated at the ATFR, the edge router cannot do the IPv4 accounting without examining the inner packet. AFTR is a 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 is only necessary if the operator requires separating per user accounting records for IPv4 and IPv6 traffic. If the per user IPv6 accounting records, collected by the edge router, are sufficient, the additional complexity of enabling IPv4 accounting at the ATFR is not required. It is important to notice that, since the IPv4 traffic is encapsulated in IPv6 packets, the data collected by the edge router for IPv6 traffic already contain the total amount of traffic (i.e. IPv4 and IPv6).

Even if detailed accounting records collection for IPv4 traffic may not be required, it would be useful for an operator in some scenarios to have, inside in the RADIUS Accounting packet, a piece of information that is generated by the BNAS for the IPv6 traffic and can be used to identify the AFTR who 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 [RFC6519].

2.9. Reliability Considerations of AFTR

The operator can use techniques such as various types of clusters to achieve high 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 with 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 on-the-
fly 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 with 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

In DS-lite, IPv4 traffic from B4 must pass through the AFTR to reach
the IPv4 Internet. Managing large numbers of tunnels and a large NAT
table could be resource intensive (e.g. CPU and memory), so the
placement of the AFTR could affect the traffic flows in the access
network and have operation implications. In general, there are two
placement models to deploy AFTR. Model One is to deploy the AFTR in
the edge of the network to cover a small region. Model Two is to
deploy the AFTR in the core of network to cover a large region.

When an operator considers where to deploy the AFTR, it 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 deploying more
AFTRs to cover the entire network. Often the operation cost
increases proportionally to the number of network equipment.

AFTR in Model Two covers a large area, thus, it serves more B4
elements. The operator could deploy only few AFTRs 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 closer
to the core network.

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. Placement of AFTR could impact the geographical aware
services. To minimize the impact, an operator could deploy AFTR
closer to users so that existing location based assumptions of the
clients source IP address by geographically aware servers can be
maintained. Another possibility is that the applications 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 handheld devices are probably usually
behind a NAT device.

2.12. Impacts on QoS

Operators commonly use DSCP [RFC2475] to classify and prioritize
different types of traffic. DS-Lite tunnel can be seen as a
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 forward 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 copies the DSCP value in the IPv4 header
to the IPv6 header after the encapsulation for the downstream traffic
and similarly the B4 copies 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 the B4 to accept incoming requests. When
the remote host is on IPv4, the incoming request will be directed
towards the B4's IPv4 address. Some applications use UPnP-IGD (e.g.,
popular gaming consoles) or ICE [RFC5245] (e.g., SIP, Yahoo!, Google,
Microsoft chat networks) to request incoming ports. Some
applications rely on ALGs or manual port configuration to reserve a
port in the NAT. In usual DS-Lite deployment, B4 does not own a
dedicated public IPv4 address or all the available ports, so it must
coordinate with its serving AFTR and the applications to reserve the
incoming ports. Port Control Protocol (PCP) [I-D.ietf-pcp-base] is
designed to address this issue.

2.14. DS-Lite Tunnel Security

Section 11 of [RFC6333] describes security issues associated to DS-
Lite mechanism. To restrict the service offered by AFTR only to
registered customers, an operator can implement IPv6 ingress filter
on the AFTR's tunnel interface to accept only the IPv6 prefixes
defined in the filter. This approach requires knowing in advance the
IPv6 prefixes provisioned to the customers for the softwire in order
to configure the filter.

Using DHCPv6 Leasequery defined in [RFC5007] is another option of
achieving the same goal and providing some form of access control to
AFTR. When the AFTR receives a packet from an unknown IPv6 prefix,
it issues a DHCPv6 Leasequery based on the DUID to the DHCPv6 server
in order to verify if that prefix was previously provisioned by the
DHCPv6 server to the specific DUID. If known, the DHCPv6 server will
reply with the IPv6 prefix and the associated lease. If both
prefixes match, the ATFR accepts the packet otherwise it drops the
packet and denies the service.

2.15. IPv6-only Network Considerations

In environments where the operator wants to deploy AFTR in the IPv6-
only network, the AFTR nodes may not have direct IPv4 connectivity.
In this scenario the operator 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 is
located in the IPv6-only network closer to B4 elements. 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 achieve 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, DHCPv6 option to RADIUS. If an
operator uses DHCPv6 to provision the B4, the B4 element must
implement the DHCPv6 option defined in [RFC6334]. If an operator
uses RADIUS to provision the B4, the B4 element must implement
[RFC6519].

3.1. DNS deployment Considerations

   [RFC6333] recommends the B4 element should send DNS queries to an
   external recursive resolver over IPv6. The B4 element should
   implement proxy resolver that will proxy DNS query from IPv4
   transport to the DNS server in the IPv6 network. Alternatively, the
   DHCPv4 server on the B4 is configured to give its clients an IPv4
   address of an external DNS recursive resolver. Then, the B4 can be
   statically configured to tunnel all DNS packets to the external
   resolver over IPv6 to the AFTR. Note that there is no effective way
   to provision an IPv4 DNS address to the B4 over IPv6, this may create
   complexity in B4 provisioning. Moreover, this will increase load to
   AFTR by creating short-live entries in the NAT table, this alternate
   solution is likely to be unsatisfactory in a production environment.
   It should be used in a testing or demonstration environment.

3.2. IPv4 Service Monitoring

3.2.1. B4 Remote Management

   B4 is connected to IPv6 access network to offer IPv4 services. When
   users experience IPv4 connectivity issue, operators must be able to
   remotely access the B4 element to verify its B4's configuration and
   status. Operators should access B4 elements using native IPv6.
   Operators should not access B4 over the softwire.

3.2.2. IPv4 Connectivity Check

   DS-Lite framework provides IPv4 services over IPv6 access network.
   Operators must be able to check the IPv4 connectivity from the B4
   element to its AFTR. AFTR should be configured with an IPv4 address
   to enable PING test and traceroute test. An operator may assign the
   same IPv4 address (e.g. 192.0.0.2/32) to all AFTRs. This IPv4
   address only used to respond to the requests from the B4 elements
   over the softwire. IANA allocates 192.0.0.0/29 [RFC6333] which can
   be used for this purpose.

4. Security Considerations

   This document does not present any new security issues. [RFC6333]
   discusses DS-Lite related security issues. General NAT security
   issues are not repeated here.

   Some of the security issues result directly from sharing routable
   IPv4 addresses. 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 customer's 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 the
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 numbers, 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,
operators 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 public 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 operators and network designers. Deployment
considerations of the B4, AFTR and DNS have been discussed and
recommendations for their usage have been documented.

6. **Acknowledgement**

Thanks to Nejc Skoberne and Maoke Chen for their through review and
helpful comments.

7. **IANA Considerations**

This memo includes no request to IANA.

8. **References**

8.1. **Normative References**

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

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

   [RFC6334]   Hankins, D. and T. Mrugalski, "Dynamic Host Configuration
               Protocol for IPv6 (DHCPv6) Option for Dual-Stack Lite",
               RFC 6334, August 2011.

   [RFC6519]   Maglione, R. and A. Durand, "RADIUS Extensions for Dual-
               Stack Lite", RFC 6519, February 2012.

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-04 (work in
               progress), September 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-pcp-base]
               Cheshire, S., Boucadair, M., Selkirk, P., Wing, D., and R.
               Penno, "Port Control Protocol (PCP)",
               draft-ietf-pcp-base-23 (work in progress), February 2012.

   [RFC1631]   Egevang, K. and P. Francis, "The IP Network Address
               Translator (NAT)", RFC 1631, May 1994.

   [RFC2473]   Conta, A. and S. Deering, "Generic Packet Tunneling in
               IPv6 Specification", RFC 2473, December 1998.

   [RFC2475]   Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z.,
               and W. Weiss, "An Architecture for Differentiated
               Services", RFC 2475, December 1998.

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

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

   [RFC5245]   Rosenberg, J., "Interactive Connectivity Establishment
               (ICE): A Protocol for Network Address Translator (NAT)
               Traversal for Offer/Answer Protocols", RFC 5245,
               April 2010.

   [RFC6269]   Ford, M., Boucadair, M., Durand, A., Levis, P., and P.
               Roberts, "Issues with IP Address Sharing", RFC 6269,
               June 2011.

   [RFC6302]   Durand, A., Gashinsky, I., Lee, D., and S. Sheppard,
               "Logging Recommendations for Internet-Facing Servers",
               BCP 162, RFC 6302, June 2011.

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
   U.S.A.

   Email: carlw@mcsr-labs.org

Christian Jacquenet
France Telecom
Rennes
France

Email: christian.jacquenet@orange.com

Mohamed Boucadair
France Telecom
Rennes
France

Email: mohamed.boucadair@orange.com