Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2011

F. Brockners
S. Gundavelli
Cisco
S. Speicher
Deutsche Telekom AG
D. Ward
Juniper Networks
October 25, 2010

Gateway Initiated Dual-Stack Lite Deployment
draft-ietf-softwire-gateway-init-ds-lite-02

Abstract

Gateway-Initiated Dual-Stack lite (GI-DS-lite) is a variant of Dual-Stack lite (DS-lite) applicable to certain tunnel-based access architectures. GI-DS-lite extends existing access tunnels beyond the access gateway to an IPv4-IPv4 NAT using softwires with an embedded context identifier that uniquely identifies the end-system the tunneled packets belong to. The access gateway determines which portion of the traffic requires NAT using local policies and sends/receives this portion to/from this softwire.

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 28, 2011.

Copyright Notice

Table of Contents

1. Overview

Gateway-Initiated Dual-Stack lite (GI-DS-lite) is a variant of the Dual-Stack lite (DS-lite) [I-D.ietf-softwire-dual-stack-lite], applicable to network architectures which use point to point tunnels between the access device and the access gateway. The access gateway in these models is designed to serve large numbers of access devices. Mobile architectures based on Mobile IPv6 [RFC3775], Proxy Mobile IPv6 [RFC5213], or GTP [TS29060], as well as broadband architectures based on PPP or point-to-point VLANs as defined by the Broadband Forum (see [TR59] and [TR101]) are examples for this type of architecture.

The DS-lite approach leverages IPv4-in-IPv6 tunnels (or other tunneling modes) for carrying the IPv4 traffic from the customer network to the Address Family Transition Router (AFTR). An established softwire between the AFTR and the access device is used for traffic forwarding purposes. This turns the inner IPv4 address irrelevant for traffic routing and allows sharing private IPv4 addresses [RFC1918] between customer sites within the service provider network.

Similar to DS-lite, GI-DS-lite enables the service provider to share public IPv4 addresses among different customers by combining tunneling and NAT. It allows multiple access devices behind the access gateway to share the same private IPv4 address [RFC1918]. Rather than initiating the tunnel right on the access device, GI-DS-lite logically extends the already existing access tunnels beyond the access gateway towards the IPv4-IPv4 NAT using a tunneling mechanism with semantics for carrying context state related to the encapsulated traffic. This approach results in supporting overlapping IPv4 addresses in the access network, requiring no changes to either the access device, or to the access architecture. Additional tunneling overhead in the access network is also omitted. If e.g., a GRE based encapsulation mechanisms is chosen, it allows the network between the access gateway and the NAT to be either IPv4 or IPv6 and provides the operator to migrate to IPv6 in incremental steps.

2. Conventions

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].

The following abbreviations are used within this document:

AFTR: Address Family Transition Router (also known as "Large Scale NAT (LSN)" or "Dual-Stack lite Tunnel Concentrator", or "Carrier Grade NAT"). An AFTR combines IP-in-IP tunnel termination and IPv4-IPv4 NAT.

AD: Access Device. It is the end host, also known as the mobile node in mobile architectures.

CID: Context Identifier

DS-lite: Dual-stack lite

GI-DS-lite: Gateway-initiated DS-lite

NAT: Network Address Translator

SW: Softwire (see [RFC4925])

SWID: Softwire Identifier

TID: Access Tunnel Identifier. The interface identifier of the point-to-point access tunnel.

3. Gateway Initiated DS-Lite

The section provides an overview of Gateway Initiated DS-Lite (GI-DS-lite). Figure 1 outlines the generic deployment scenario for GI-DS-lite. This generic scenario can be mapped to multiple different access architectures, some of which are described in Section 7.

In Figure 1, access devices (AD-1 and AD-2) are connected to the Gateway using some form of tunnel technology and the same is used for carrying IPv4 (and optionally IPv6) traffic of the access device. These access devices may also be connected to the Gateway over point-to-point links. The details on how the network delivers the IPv4 address configuration to the access devices are specific to the access architecture and are outside the scope of this document. With GI-DS-lite, Gateway and AFTR are connected by a softwire [RFC4925]. The softwire is identified by a softwire identifier (SWID). The form of the SWID depends on the tunneling technology used for the softwire. The SWID could e.g. be the endpoints of a GRE-tunnel or a VPN-ID, see Section 6 for details. A Context-Identifier (CID) is used to multiplex flows associated with the individual access devices onto the softwire. Local policies at the Gateway determine which part of the traffic received from an access device is tunneled over the softwire to the AFTR. The combination of CID and SWID (potentially along with other traffic identifiers such as e.g.

interface, VLAN, port, etc.) serves as common context between Gateway and AFTR to uniquely identify flows associated with an access device. The CID is typically a 32-bit wide identifier and is assigned by the Gateway. It is retrieved either from a local or remote (e.g. AAA) repository. Like the SWID, the embodiment of the CID depends on the tunnel mode used and the type of the network connecting Gateway and AFTR. If, for example GRE [RFC2784] with "GRE Key and Sequence Number Extensions" [RFC2890] is used as softwire technology, the network connecting Gateway and AFTR could be either IPv4-only, IPv6-only, or a dual-stack IP network. The CID would be carried within the GRE-key field. See Section 6 for details on different softwire types supported with GI-DS-lite.

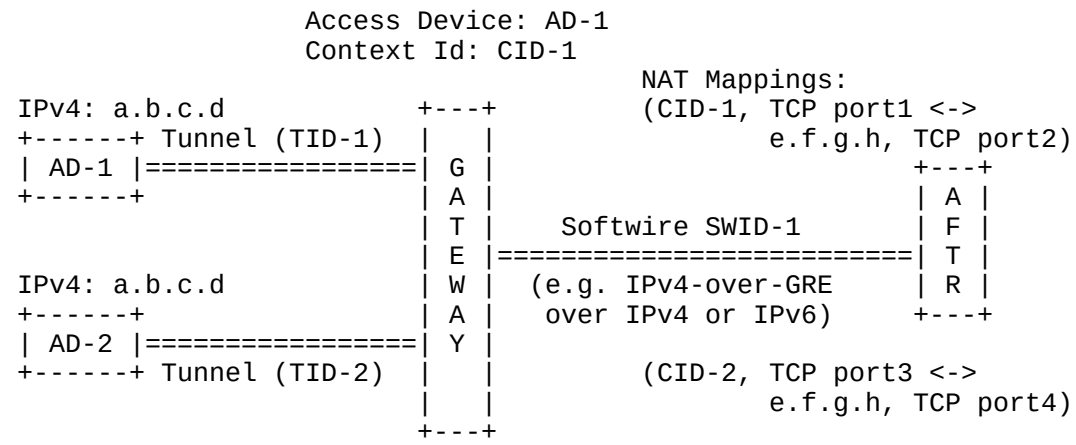

Figure 1: Gateway-initiated dual-stack lite reference architecture

The AFTR combines softwire termination and IPv4-IPv4 NAT. The outer/external IPv4 address of a NAT-binding at the AFTR is either assigned autonomously by the AFTR from a local address pool, configured on a per-binding basis (either by a remote control entity through a NAT control protocol or through manual configuration), or derived from the CID (e.g., the CID, in case 32-bit wide, could be mapped 1:1 to an external IPv4-address). A simple example of a translation table at the AFTR is shown in Figure 2. The choice of the appropriate translation scheme for a traffic flow can take parameters such as destination IP-address, incoming interface, etc. into account. The IP-address of the AFTR, which, depending on the transport network between the Gateway and the AFTR, will either be an IPv6 or an IPv4 address, is configured on the Gateway. A variety of methods, such as

out-of-band mechanisms, or manual configuration apply.

Softwire-Id/Context-Id/IPv4/Port	Public IPv4/Port
SWID-1/CID-1/a.b.c.d/TCP-port1	e.f.g.h/TCP-port2
SWID-1/CID-2/a.b.c.d/TCP-port3	e.f.g.h/TCP-port4

Figure 2: Example translation table on the AFTR

GI-DS-lite does not require a 1:1 relationship between Gateway and AFTR, but more generally applies to (M:N) scenarios, where M Gateways are connected to N AFTRs. Multiple Gateways could be served by a single AFTR. AFTRs could be dedicated to specifc groups of access-devices, groups of Gateways, or geographic regions. An AFTR could, but does not have to be co-located with a Gateway.

4. Protocol and related Considerations

- The NAT binding entry maintained at the AFTR, which reflects an active flow between an access device inside the network and a node in the Internet, needs to be extended to include two other parameters, the CID and the identifier of the softwire (SWID).

- When creating an IPv4 to IPv4 NAT binding for an IPv4 packet flow received from the Gateway over the softwire, the AFTR will associate the CID with that NAT binding. It will use the combination of CID and SWID as the unique identifier and will store it in the NAT binding entry.

- When forwarding a packet to the access device, the AFTR will obtain the CID from the NAT binding associated with that flow. E.g., in case of GRE-encapsulation, it will add the CID to the GRE Key and Sequence number extension of the GRE header and tunnel it to the Gateway.

- On receiving any packet from the softwire, the AFTR will obtain the CID from the incoming packet and will use it for performing the NAT binding look up and for performing the packet translation before forwarding the packet.

- The Gateway, on receiving any IPv4 packet from the access device will lookup the CID for that access device. In case of GRE

encapsulation it will for example add the CID to the GRE Key and Sequence number extension of the GRE header and tunnel it to the AFTR.

- On receiving any packet from the softwire, the Gateway will obtain the CID from the packet and will use it for making the forwarding decision. There will be an association between the CID and the forwarding state.

- When encapsulating and IPv4 packet, Gateway and AFTR can its Diffserv Codepoint (DSCP) to derive the DSCP (or MPLS Traffic-Class Field in case of MPLS) of the softwire.

5. Softwire Management and related Considerations

The following are the considerations related to the operational management of the softwire between AFTR and Gateway.

- The softwire between the Gateway and the AFTR is created at system startup time and stays up active all time. Deployment dependent, Gateway and AFTR can employ OAM mechanisms such as ICMP, BFD [RFC5880], or LSP ping [RFC4379] for softwire health management and corresponding protection strategies.

- The softwire peers may be provisioned to perform policy enforcement, such as for determining the protocol-type or overall portion of traffic that gets tunneled, or for any other quality of service related settings. The specific details on how this is achieved or the types of policies that can be applied are outside the scope for this document.

- The softwire peers must have a proper understanding of the path MTU value. This can be statically configured at softwire creation time.

- A Gateway and an AFTR can have multiple softwires established between them (e.g. to separate address domains, provide for load-sharing etc.).

6. Softwire Embodiments

Deployment and requirements dependent, different tunnel technologies apply for the softwire connecting Gateway and AFTR. GRE encapsulation with GRE-key extensions, MPLS VPNs, or plain IP-in-IP encapsulation can be used. Softwire identification and Context-ID depend on the tunneling technology employed:

- GRE with GRE-key extensions: Softwire identification is supplied by the endpoints of the GRE tunnel. The GRE-key serves as CID.

- MPLS VPN: Softwire identification is supplied by the VPN identifier of the MPLS VPN. The IPv4-address serves as CID. The IPv4-address within a VPN has to be unique.

- Plain IP-in-IP: Softwire identification is supplied by the endpoints of the IP-in-IP tunnel. Either the inner IPv4-address serves as CID (in which case the IPv4-address has to be unique) or the IPv6-Flow-Label serves as CID (which obviously only applies to cases where IPv6 transport is used). Note that when using the IP-Flow-Label as CID additional scaling considerations might apply given that the CID is to only 20 bits wide in this case. Also one should ensure sufficient randomization in this case to for example avoid interference with other uses of the IP-Flow-Label, such as ECMP.

Figure 3 gives an overview of the different tunnel modes as they apply to different deployment scenarios. "x" indicates that a certain deployment scenario is supported. The following abbreviations are used:

- IPv4 address

  * "up": Deployments with "unique private IPv4 addresses" assigned to the access devices are supported.

  * "op": Deployments with "overlapping private IPv4 addresses" assigned to the access devices are supported.

  * "nm": Deployments with "non-meaningful/dummy but unique IPv4 addresses" assigned to the access devices are supported.

  * "s": Deployments where all access devices are assigned the same IPv4 address are supported.

- Network-type

  * "v4": Gateway and AFTR are connected by an IPv4-only network

  * "v6": Gateway and AFTR are connected by an IPv6-only network

  * "v4v6": Gateway and AFTR are connected by a dual stack network, supporting IPv4 and IPv4.

  * "MPLS": Gateway and AFTR are connected by a MPLS network

Softwire	IPv4 address				Network-type			
	up	op	nm	s	v4	v6	v4v6	MPLS
GRE with GRE-key	x	x	x	x	x	x	x	
MPLS VPN	x	x	x					x
Plain IP-in-IP	x	x	x	x	x	x	x	

Figure 3: Tunnel modes and their applicability

Note: For "Plain IP-in-IP", support for ’op’ and ’s’ requires the use of IPv6-transport with the IPv6-Flow-Label serving as CID.

7. GI-DS-lite deployment

7.1. Connectivity establishment: Example call flow

Figure 4 shows an example call flow - linking access tunnel establishment on the Gateway with the softwire to the AFTR. This simple example assumes that traffic from the AD uses a single access tunnel and that the Gateway will use local polices to decide which portion of the traffic received over this access tunnel needs to be forwarded to the AFTR.

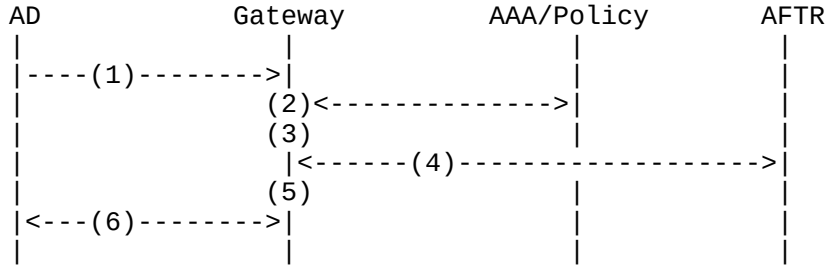

Figure 4: Example call flow for session establishment

1. Gateway receives a request to create an access tunnel endpoint.

2. The Gateway authenticates and authorizes the access tunnel. Based on local policy or through interaction with the AAA/Policy system the Gateway recognizes that IPv4 service should be provided using GI-DS-lite.

3. The Gateway creates an access tunnel endpoint. The access tunnel links AD and Gateway and is uniquely identified by Tunnel Identifier (TID) on the Gateway.

4. (Optional): The Gateway and the AFTR establish a control session between each other. This session can for example be used to exchange accounting or NAT-configuration information. Accounting information could be supplied to the Gateway, AAA/Policy, or other network entities which require information about the externally visible address/port pairs of a particular access device. The Diameter NAT Control Application (see [I-D.draft-ietf-dime-nat-control] could for example be used for this purpose.

5. The Gateway allocates a unique CID and associates those flows received from the access tunnel (identified by the TID) that need to be tunneled towards the AFTR with the softwire linking Gateway and AFTR. Local forwarding policy on the Gateway determines which traffic will need to be tunneled towards the AFTR.

6. Gateway and AD complete the access tunnel establishment (depending on the procedures and mechanisms of the corresponding access network architecture this step can include the assignment of an IPv4 address to the AD).

7.2. GI-DS-lite applicability: Examples

The section outlines deployment examples of the generic GI-DS-lite architecture described in Section 3.

- Mobile IP based access architectures: In a MIPv6 [RFC5555] based network scenario, the Mobile IPv6 home agent will implement the GI-DS-lite Gateway function along with the dual-stack Mobile IPv6 functionality.

- Proxy Mobile IP based access architectures: In a PMIPv6 [RFC5213] scenario the local mobility anchor (LMA) will implement the GI-DS-lite Gateway function along with the PMIPv6 IPv4 support functionality.

- GTP based access architectures: 3GPP TS 23.401 [TS23401] and 3GPP TS 23.060 [TS23060] define mobile access architectures using GTP. For GI-DS-lite, the PDN-Gateway/GGSN will also assume the Gateway function.

- Fixed WiMAX architecture: If GI-DS-lite is applied to fixed WiMAX, the ASN-Gateway will implement the GI-DS-lite Gateway function.

- Mobile WiMAX: If GI-DS-lite is applied to mobile WiMAX, the home agent will implement the Gateway function.

- PPP-based broadband access architectures: If GI-DS-lite is applied to PPP-based access architectures the Broadband Remote Access Server (BRAS) or Broadband Network Gateway (BNG) will implement the GI-DS-lite Gateway function.

- In broadband access architectures using per-subscriber VLANs the BNG will implement the GI-DS-lite Gateway function.

8. Acknowledgements

The authors would like to acknowledge the discussions on this topic with Mark Grayson, Jay Iyer, Kent Leung, Vojislav Vucetic, Flemming Andreasen, Dan Wing, Jouni Korhonen, Teemu Savolainen, Parviz Yegani, Farooq Bari, Mohamed Boucadair, Vinod Pandey, Jari Arkko, Eric Voit, Yiu L. Lee, Tina Tsou, Guo-Liang Yang, and Cathy Zhou.

9. IANA Considerations

This document includes no request to IANA.

All drafts are required to have an IANA considerations section (see the update of RFC 2434 [RFC5226] for a guide). If the draft does not require IANA to do anything, the section contains an explicit statement that this is the case (as above). If there are no requirements for IANA, the section will be removed during conversion into an RFC by the RFC Editor.

10. Security Considerations

All the security considerations from GTP [TS29060], Mobile IPv6 [RFC3775], Proxy Mobile IPv6 [RFC5213], and Dual-Stack lite [I-D.ietf-softwire-dual-stack-lite] apply to this specification as well.

11. Change History (to be removed prior to publication as an RFC)

Changes from -00 to -01

a. clarified the applicability of GI-DS-lite to scenarios with M Gateways and N AFTRs.

b. clarification of the nomenclature and use of the identifier of the softwire connecting Gateway and AFTR: Introduced softwire identifier (SWID), updated figure 2 accordingly.

c. cleanup of editorial nits.

d. added IP-Flow-Label as CID.

Changes from -00 to -02

a. added considerations for the use of the IP-Flow-Label as CID.

b. editorial edits (additional acknowledgements).

12. References

12.1. 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.

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, March 2000.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, September 2000.

[RFC3775] Johnson, D., Perkins, C., and J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.

[RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.

[RFC5213] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC5555] Soliman, H., "Mobile IPv6 Support for Dual Stack Hosts and Routers", RFC 5555, June 2009.

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

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.

12.2. Informative References

[I-D.draft-ietf-dime-nat-control] Brockners, F., Bhandari, S., Singh, V., and V. Fajardo, "Diameter NAT Control Application", August 2009.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", RFC 3032, January 2001.

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

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

[TR101] Broadband Forum, "TR-101: Migration to Ethernet-Based DSL Aggregation", April 2006.

[TR59] Broadband Forum, "TR-059: DSL Evolution - Architecture Requirements for the Support of QoS-Enabled IP Services", September 2003.

[TS23060] "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS); Service description; Stage 2.", 2009.

[TS23401] "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN)

access.", 2009.

[TS29060] "3rd Generation Partnership Project; Technical Specification Group Core Network and Terminals; General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP), V9.1.0", 2009.

Authors' Addresses

Frank Brockners
Cisco
Hansaallee 249, 3rd Floor
DUESSELDORF, NORDRHEIN-WESTFALEN 40549
Germany

Email: fbrockne@cisco.com

Sri Gundavelli
Cisco
170 West Tasman Drive
SAN JOSE, CA 95134
USA

Email: sgundave@cisco.com

Sebastian Speicher
Deutsche Telekom AG
Landgrabenweg 151
BONN, NORDRHEIN-WESTFALEN 53277
Germany

Email: sebastian.speicher@telekom.de

David Ward
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, California 94089-1206
USA

Email: dward@juniper.net