

SOFTWARE WG  
Internet-Draft  
Intended status: Standards Track  
Expires: October 30, 2012

F. Brockners  
S. Gundavelli  
Cisco  
S. Speicher  
Deutsche Telekom AG  
D. Ward  
Cisco  
April 28, 2012

Gateway Initiated Dual-Stack Lite Deployment  
draft-ietf-softwire-gateway-init-ds-lite-08

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 October 30, 2012.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Overview . . . . .	3
2. Conventions . . . . .	3
3. Gateway Initiated DS-Lite . . . . .	4
4. Protocol and related Considerations . . . . .	6
5. Software Management and related Considerations . . . . .	7
6. Software Embodiments . . . . .	7
7. IANA Considerations . . . . .	9
8. Security Considerations . . . . .	9
9. Acknowledgements . . . . .	10
10. References . . . . .	10
10.1. Normative References . . . . .	10
10.2. Informative References . . . . .	11
Appendix A. GI-DS-lite deployment . . . . .	12
A.1. Connectivity establishment: Example call flow . . . . .	12
A.2. GI-DS-lite applicability: Examples . . . . .	13
Authors' Addresses . . . . .	14

## 1. Overview

Gateway-Initiated Dual-Stack lite (GI-DS-lite) is a variant of the Dual-Stack lite (DS-lite) [RFC6333], 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 [RFC6275], 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 [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 software 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 Address Family Transition Router (AFTR) 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., GRE based encapsulation mechanisms is chosen, it allows the network between the access gateway and the AFTR 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. 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 [RFC4925]

SWID: Softwire Identifier

### 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 Appendix A.

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 SWID does not need to be globally unique, i.e. different SWIDs could be used to identify a softwire at the different ends of a softwire. 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, Section 6 for details. A Context-Identifier (CID) is used to multiplex flows associated with the individual access devices onto the softwire. Deployment dependent, the flows from a particular AD can be identified using either the source IP-address or an access tunnel identifier. 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 must be unique between gateway and AFTR to identify the flows associated with an AD. 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 software 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. Section 6 for details on different software types supported with GI-DS-lite.

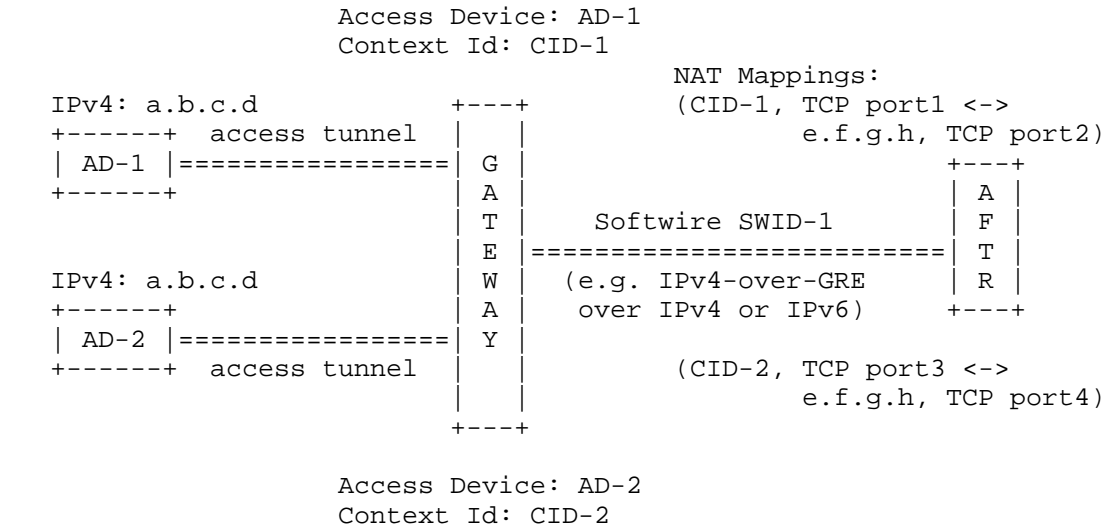


Figure 1: Gateway-initiated dual-stack lite reference architecture

The AFTR combines software termination and IPv4-IPv4 NAT. The NAT binding of the AD's address could be 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.

Software-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 specific 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

- o Depending on the embodiment of the CID (e.g. for GRE-encapsulation with GRE-key), 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, SHOULD be extended to include the CID and the identifier of the software (SWID).
- o When creating an IPv4 to IPv4 NAT binding for an IPv4 packet flow received from the Gateway over the software, the AFTR SHOULD associate the CID with that NAT binding. It SHOULD use the combination of CID and SWID as the unique identifier and use those parameters in the NAT binding entry.
- o When forwarding a packet to the access device, the AFTR SHOULD obtain the CID from the NAT binding associated with that flow. E.g., in case of GRE-encapsulation, it SHOULD add the CID to the GRE Key and Sequence number extension of the GRE header and tunnel it to the Gateway.
- o On receiving any packet from the software, the AFTR SHOULD obtain the CID from the incoming packet and use it for performing the NAT binding look up and for performing the packet translation before forwarding the packet.
- o The Gateway, on receiving any IPv4 packet from the access device SHOULD lookup the CID for that access device. In case of GRE encapsulation it can for example add the CID to the GRE Key and Sequence number extension of the GRE header and tunnel it to the

AFTR.

- o On receiving any packet from the software, the Gateway SHOULD obtain the CID from the packet and use it for making the forwarding decision.
- o When encapsulating an IPv4 packet, Gateway and AFTR SHOULD use its Diffserv Codepoint (DSCP) to derive the DSCP (or MPLS Traffic-Class Field in case of MPLS) of the software.

## 5. Software Management and related Considerations

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

- o The software between the Gateway and the AFTR MAY be created at system startup time, OR dynamically established on-demand. Deployment dependent, Gateway and AFTR can employ OAM mechanisms such as ICMP, BFD [RFC5880], or LSP ping [RFC4379] for software health management and corresponding protection strategies.
- o The software 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.
- o The software peers SHOULD use the correct path MTU value for the tunnel path between the access gateway and the AFTR. This value MAY be statically configured at software creation time, or dynamically discovered using the standard path MTU discovery techniques.
- o A Gateway and an AFTR can have multiple softwares established between them (e.g. to separate address domains, provide for load-sharing etc.).

## 6. Software Embodiments

Deployment and requirements dependent, different tunnel technologies apply for the software connecting Gateway and AFTR. GRE encapsulation with GRE-key extensions, MPLS VPNs [RFC4364], or plain IP-in-IP encapsulation can be used. Software identification and Context-ID depend on the tunneling technology employed:

- o GRE with GRE-key: SWID is the tunnel identifier of the GRE tunnel between the GW and the AFTR. The CID is the GRE-key associated with the AD.
- o MPLS VPN: The SWID is a generic identifier which uniquely identifies the VPN at either the Gateway or AFTR. Depending on whether the Gateway or AFTR are acting as customer edge (CE) or, provider edge (PE), the SWID could e.g. be an attachment circuit identifier, an identifier representing the set of VPN route labels pointing to the routes within the VPN, etc. The AD's IPv4-address is the CID. For a given VPN, the AD's IPv4 address must be unique.
- o IPv4/IPv6-in-MPLS: The SWID is the top MPLS label. CID might be the next MPLS label in the stack, if present, or the IP address of the AD.
- o IPv4-in-IPv4: SWID is the outer IPv4 source address. The AD's IPv4 address is the CID. For a given outer IPv4 source address, the AD's IPv4 address must be unique.
- o IPv4-in-IPv6: SWID is the outer IPv6 source address. If the AD's IPv4 address is used as CID, the AD's IPv4 address must be unique. If the IPv6-Flow-Label [RFC6437] is used as CID, the IPv4 addresses of the ADs may overlap. Given that the IPv6-Flow-Label is 20-bit wide, which is shorter than the recommended 32-bit CID, large scale deployments may require additional scaling considerations. In addition, one should ensure sufficient randomization of the IP-Flow-Label to avoid possible interference with other uses of the IP-Flow-Label, such as Equal Cost Multipath (ECMP) support.

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:

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



- o 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 IPv6.
- \* "MPLS": Gateway and AFTR are connected by a MPLS network

Software	IPv4 address			Network-type			
	up	op	s	v4	v6	v4v6	MPLS
GRE with GRE-key	x	x	x	x	x	x	
MPLS VPN	x	x					x
IPv4/IPv6-in-MPLS	x	x	x				x
IPv4-in-IPv4	x			x			
IPv4-in-IPv6	x				x		
IPv4-in-IPv6 w/ FL	x	x	x		x		

Figure 3: Tunnel modes and their applicability

## 7. IANA Considerations

This specification does not require any IANA actions.

## 8. Security Considerations

The approach specified in this document allows the use of Dual-stack lite for tunnel-based access architectures. Rather than initiating the tunnel from the access device, GI-DS-lite logically extends the already existing access tunnel beyond the access gateway towards the Address Family Transition Router, and builds a virtual software between the AFTR and the access device. This approach requires the use of an additional context identifier in the AFTR and at the access gateway, which is required for making IP packet forwarding decisions.

A packet when received with an Incorrect context identifier at the access gateway/AFTR will result in associating the packet to an incorrect access device. Therefore, care must be taken to ensure an IP packet tunneled between the access gateway and the AFTR is carried

with the context identifier of the access device associated with that IP packet. The context identifier is not carried from the access device and it is not possible for one access device to claim the context identifier of some other access device. However, It is possible an on-path attacker between the access gateway and the AFTR can potentially modify the context identifier in the packet, resulting in association of the packet to an incorrect access device. This threat is no different from an on-path attacker modifying the source/destination address of an IP packet. However, this threat can be prevented by enabling IPsec security with integrity protection turned on, between the access gateway and the AFTR, that will ensure the correct binding of the context identifier and the inner packet. This specification does not require any other new security considerations other than those specified in dual-stack lite specification [RFC6333], and in the security considerations specified for the given access architecture, such as Proxy Mobile IPv6, leveraging this transitioning scheme.

## 9. 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, Cathy Zhou, Olaf Bonness, Paco Cortes, Jim Guichard, Stephen Farrell, Pete Resnik, Ralph Droms.

## 10. References

### 10.1. Normative References

- [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.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private

Networks (VPNs)", RFC 4364, February 2006.

- [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.
- [RFC5555] Soliman, H., "Mobile IPv6 Support for Dual Stack Hosts and Routers", RFC 5555, June 2009.
- [RFC5844] Wakikawa, R. and S. Gundavelli, "IPv4 Support for Proxy Mobile IPv6", RFC 5844, May 2010.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.
- [RFC6275] Perkins, C., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, July 2011.
- [RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.
- [RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, November 2011.

## 10.2. Informative References

- [I-D.draft-ietf-dime-nat-control] Brockners, F., Bhandari, S., Singh, V., and V. Fajardo, "Diameter NAT Control Application", August 2009.
- [RFC4925] Li, X., Dawkins, S., Ward, D., and A. Durand, "Softwire Problem Statement", RFC 4925, July 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.

## Appendix A. GI-DS-lite deployment

### A.1. Connectivity establishment: Example call flow

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

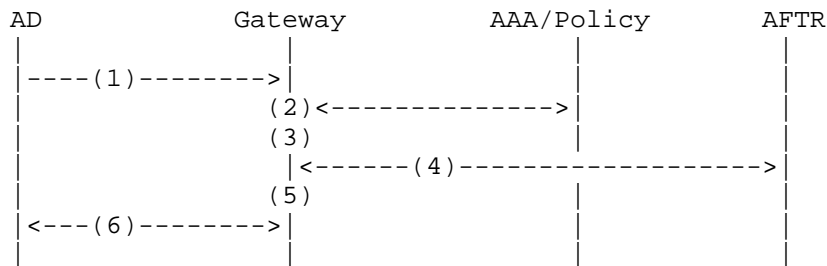


Figure 4: Example call flow for session establishment

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.
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 [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 that need to be tunneled towards the AFTR with the software 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).

#### A.2. GI-DS-lite applicability: Examples

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

- o Mobile IP based access architectures: In a DSMIPv6 [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.
- o Proxy Mobile IPv6 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 [RFC5844] functionality.
- o 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.
- o Fixed WiMAX architecture: If GI-DS-lite is applied to fixed WiMAX, the ASN-Gateway will implement the GI-DS-lite Gateway function.
- o Mobile WiMAX: If GI-DS-lite is applied to mobile WiMAX, the home agent will implement the Gateway function.
- o 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.

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

#### Authors' Addresses

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

Email: [fbrockne@cisco.com](mailto:fbrockne@cisco.com)

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

Email: [sgundave@cisco.com](mailto:sgundave@cisco.com)

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

Email: [sebastian.speicher@telekom.de](mailto:sebastian.speicher@telekom.de)

David Ward  
Cisco  
170 West Tasman Drive  
SAN JOSE, CA 95134  
USA

Email: [wardd@cisco.com](mailto:wardd@cisco.com)

