

SPRING Working Group
Internet Draft
Intended status: Standards Track
Expires: June 22, 2018

Z. Ali
C. Filsfils
R. Gandhi
N. Kumar
F. Iqbal
C. Pignataro
Cisco Systems, Inc.
D. Steinberg
Steinberg Consulting
S. Salsano
Universita di Roma "Tor Vergata"
G. Naik
Drexel University
December 23, 2017

Performance Measurement in Segment Routing Networks
draft-ali-spring-sr-pm-00.txt

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on June 22, 2018.

Copyright Notice

Copyright (c) 2017 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.

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 as described in Section 4.e of the [Trust Legal Provisions](#) and are provided without warranty as described in the Simplified BSD License.

Abstract

[RFC 6374](#) specifies protocol mechanisms to enable the efficient and accurate measurement of packet loss, one-way and two-way delay, as well as related metrics such as delay variation and channel throughput in MPLS networks. This document describes how these mechanisms can be used for performance measurements in Segment Routing with MPLS data plane (SR-MPLS) networks. The document also specifies how similar mechanisms can be used for performance measurement in Segment Routing over IPv6 data plane (SRv6) networks.

Table of Contents

1.	Introduction.....	2
2.	Performance Measurement in SR-MPLS Networks.....	3
2.1.	Delay Measurement in SR-MPLS Networks.....	3
2.1.1.	Delay Measurement Message Format.....	3
2.1.2.	One Way Delay Measurement.....	3
2.1.2.1.	One-Way Delay Measurement using Synthetic Probes	3
2.1.3.	Two Way Delay Measurement.....	3
3.	Performance Measurement in SRv6 Networks.....	4
3.1.	Terminology and Reference Topology.....	4
3.2.	Delay Measurement in SRv6 Networks.....	5
3.2.1.	One Way Delay Measurement.....	5
3.2.2.	Two Way Delay Measurement.....	6
3.2.3.	Delay Measurement Message Format.....	6
3.2.4.	One-Way Delay Measurement using Synthetic Probes.....	8
3.2.4.1.	Example Procedure.....	8
3.2.5.	In-situ One-Way Segment-by-Segment Delay Measurement	9
3.2.5.1.	Example Procedure.....	9
4.	Security Considerations.....	10
5.	IANA Considerations.....	10
6.	Contributors.....	10
7.	References.....	10
7.1.	Normative References.....	10
7.2.	Informative References.....	11
8.	Acknowledgments.....	11

1. Introduction

Service provider's ability to satisfy Service level agreements (SLAs) depend on the ability to measure and monitor performance metrics for packet loss and one-way and two-way delay, as well as related metrics such as delay variation and channel throughput. The ability to monitor these performance metrics also provides network operators with greater visibility into the performance characteristics of their networks.

ali, et al.

Expires June 23, 2018

[Page 2]

Internet-Draft

SR Performance Measurement

December 2017

networks, thereby facilitating planning, troubleshooting, and network performance evaluation. [RFC6374] specifies protocol mechanisms to enable the efficient and accurate measurement of these performance metrics in MPLS networks. This document describes how these mechanisms can be used for performance measurements in Segment Routing with the MPLS data plane (SR-MPLS) networks. The document also specifies how similar mechanisms can be used for performance measurement in Segment Routing with the IPv6 data plane (SRv6) networks.

2. Performance Measurement in SR-MPLS Networks

SR-MPLS relies on MPLS data plane without any changes. Hence, the protocol mechanisms for MPLS networks defined in [RFC6374] are equally applicable to SR-MPLS networks. This version of the document focuses on delay measurement. Measurements for loss and other performance metrics are to be added in the future version of this document.

2.1. Delay Measurement in SR-MPLS Networks

2.1.1. Delay Measurement Message Format

As described in [RFC6374], Section 2.9.1, MPLS DM probe messages flow over the MPLS Generic Associated Channel (G-ACh). Thus, a probe packet for a DM message contains SR-MPLS label stack, with the G-ACh Label (GAL) at the bottom of the stack. The GAL is followed by an Associated Channel Header (ACH) (value 0x00 for delay measurement) [RFC6374], which identifies the message type, and the message body following the ACH. The format of the DM message payload as defined in [RFC6374] is used for SR-MPLS delay measurement.

2.1.2. One Way Delay Measurement

2.1.2.1. One-Way Delay Measurement using Synthetic Probes

The query and response mechanisms defined in [RFC6374] are followed for synthetic delay measurement in SR-MPLS network.

For one-way delay measurement, the querier node SHOULD send the UDP Return 0 (URO) (Type=131) defined in [RFC7867]. The responder node SHOULD send the response back to the querier node in an UDP message when the URO TLV is present in the PM query message.

2.1.3. Two Way Delay Measurement

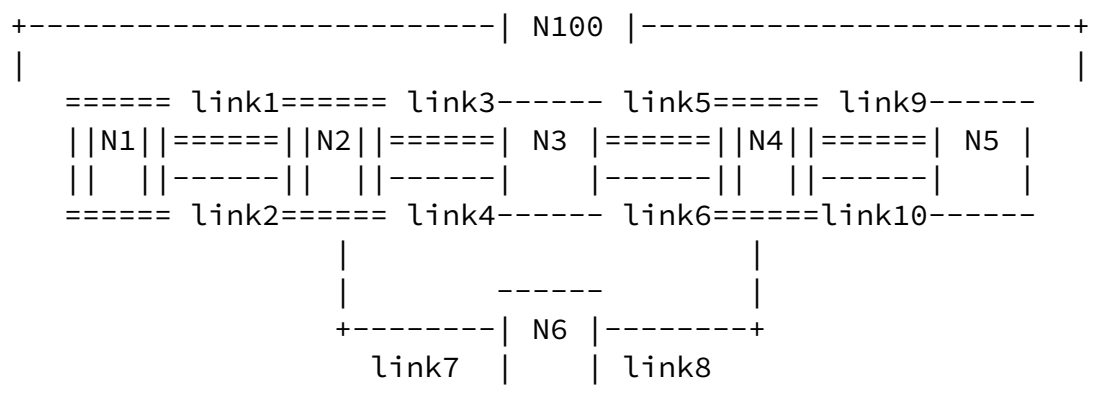
The two-way delay measurement for packet networks is defined in [RFC6374]. The two-way delay measurement in SR-MPLS networks is to be added in the future version of this document.

3. Performance Measurement in SRv6 Networks

This version of the document focuses on delay measurement. Loss and other performance metric measurements are to be added in the future version of this document.

3.1. Terminology and Reference Topology

Throughout the document, the following simple topology is used for illustration.



Reference Topology

In the reference topology:

Nodes N1, N2, and N4 are SRv6 capable nodes.

Nodes N3, N5 and N6 are classic IPv6 nodes.

Node 100 is a controller.

Node N_k has a classic IPv6 loopback address B_k::/128

Node N_k has A_k::/48 for its local SID space from which Local SIDs are explicitly allocated.

The IPv6 address of the *n*th Link between node X and Y at the X side is represented as 99:X:Y::X_n. e.g., the IPv6 address of link6 (the 2nd link) between N3 and N3 in Figure 1 is 99:3:4:32. Similarly, the IPv6 address of link5 (the 1st link between N3 and N4) at node 3 is 99:3:4::31.

A_k::0 is explicitly allocated as the END function at Node k.

A_k::C_{ij} is explicitly allocated as the END.X function at node k towards neighbor node i via *j*th Link between node i and node j. e.g., A2::C31 represents END.

N2 towards N3 via link3 (the 1st link between N2 and N3). Similarly, A4::C5 represents the END.X at N4 towards N5 via link10.

SRH is the abbreviation for the Segment Routing Header.

SL is the abbreviation for the Segment Left.

SID is the abbreviation for the Segment ID.

<S1, S2, S3> represents a SID list where S1 is the first SID and S3 is the last SID. (S3, S2, S1; SL) represents the same SID list but encoded in the SRH format where the rightmost SID (S1) in the SRH is the first SID and the leftmost SID in the SRH is the last SID.

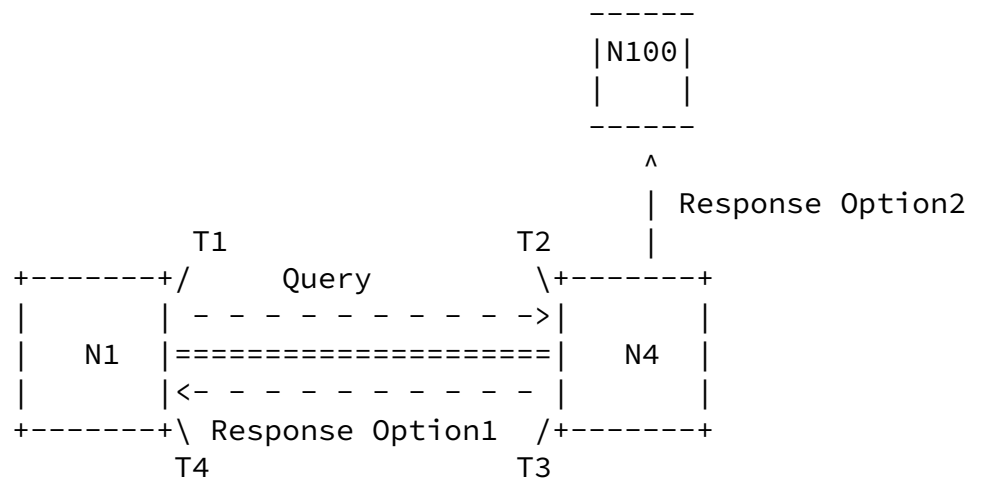
ECMP is the abbreviation for the Equal Cost Multi-Path.

UCMP is the abbreviation for the Unequal Cost Multi-Path.

[3.2.](#) Delay Measurement in SRv6 Networks

[3.2.1.](#) One Way Delay Measurement

The one-way delay measurement for packet networks is defined in [\[RFC2679\]](#), further exemplified using the following Figure.



Delay Measurement Reference Model

Nodes N1 and N4 may not be directly connected, as shown in the reference topology in Figure 1. When N1 and N4 are not directly connected, the one-way delay measurement reflects the delay observed by the packet over an arbitrary SRv6 segment-list/ policy. In other words, the one-way delay is associated with forward (N1 to N4) direction of the SRv6 segment-list/ policy.

The delay measurement can be performed using Active (using synthetic probe) and Passive (using data stream aka in-situ) mode. In both modes, T1 refers to the time the packet is transmitted from N1. Timestamping is done as late as possible at the egress pipeline (in hardware) at node N1. T2 refers to the time the packet is received at N2. Timestamping at the receiver (N2) is done as soon as possible at the ingress pipeline (in hardware).

The one-way delay metric can be defined as follows [RFC2679], [RFC6374],

$$\text{One-way delay} = T2 - T1.$$

Clock synchronization using methods detailed in [RFC6374] is assumed here.

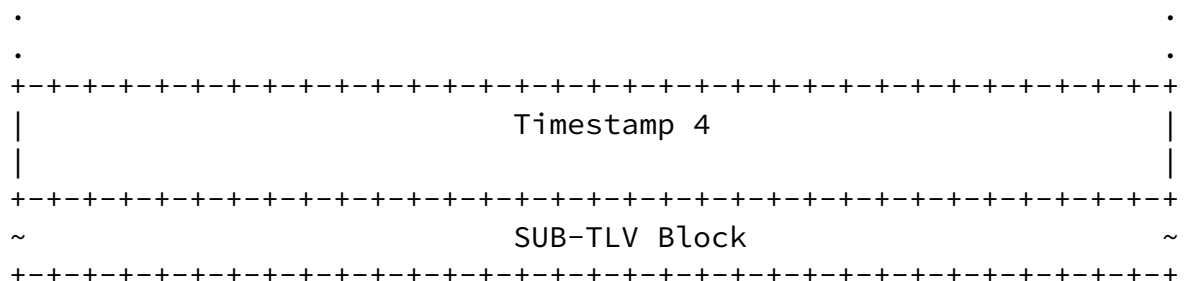
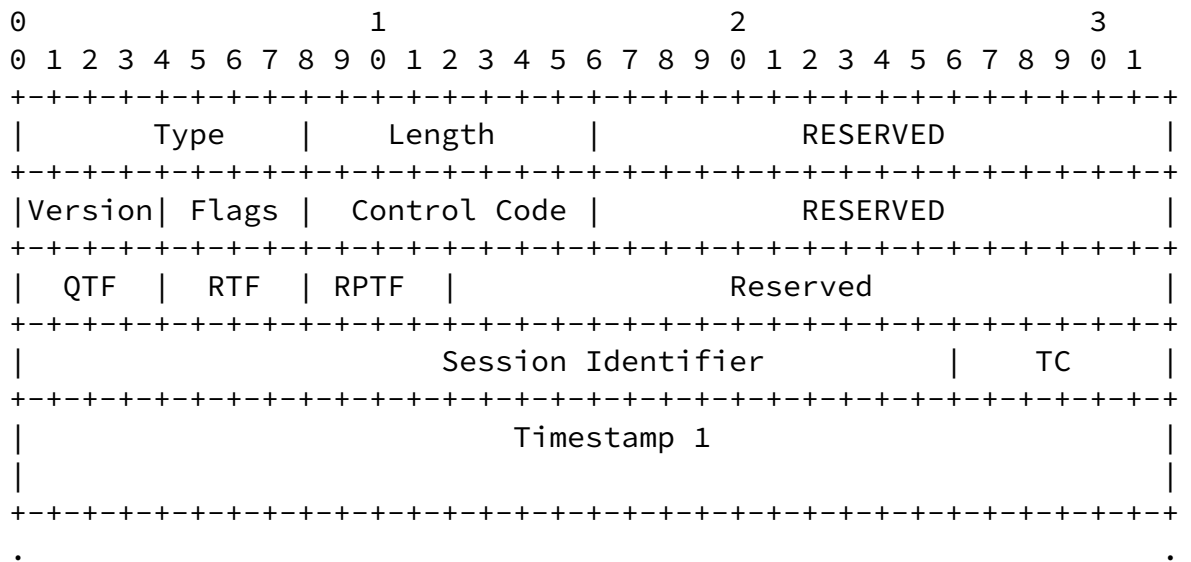
Please note that for the one-way delay computation, the receiver (node N4 in Figure 2) is not required to send a response. The response can be sent to a controller (node N100 in Figure 2). The controller may also request the querier (node N1 in Figure 2) to initiate a measurement (this messaging is not shown in Figure 2 and is beyond the scope of this document).

3.2.2. Two Way Delay Measurement

The two-way delay measurement for packet networks is defined in [RFC6374]. two-way delay measurement in SRv6 networks is to be added in the future vers this document.

3.2.3. Delay Measurement Message Format

[I-D.[draft-ietf-6man-segment-routing-header](#)] defines Segment Routing Header for SRv6. SRH can contain TLVs, as specified in [I-D.[draft-ietf-6man-segment-routing-header](#)]. This document specifies Delay Measurement (DM) TLV of SRH. DM TLV adapts a message format similar to the message format specified in [RFC6374]. The DM TLV format in SRv6 network is defined as following:



The meanings of the fields are summarized in the following table.

Field	Meaning
-------	---------

Type	SRH TLV type (Value TBA)
Length	Total length of the TLV in bytes
Version	Protocol version
Flags	Message control flags
Control Code	Code identifying the query or response type
QTF	Querier timestamp format
RTF	Responder timestamp format
RPTF	Responder's preferred timestamp format
Reserved	Reserved for future specification
Session Identifier	Set arbitrarily by the querier
Traffic Class (TC) Field	Traffic Class being measured
Timestamp 1-4	64-bit timestamp values (as shown in Figure 2)
TLV Block	Optional block of Type-Length-Value fields

Reserved fields MUST be set to 0 and ignored upon receipt. The possible values for the remaining fields are as follows.

Version: Currently set to 1 (to identify definition of TC field in [\[RFC6374\]](#))

Flags: As specified in [\[RFC6374\]](#). The T flag in a DM message is set to 1.

Control Code: As specified in [\[RFC6374\]](#).

Message Length: Set to the total length of this message in bytes, including Version, Flags, Control Code, and Message Length fields as well as the TLV Block if any.

Querier Timestamp Format: The format of the timestamp values written by the querier, as specified in [Section 3.4 of \[RFC6374\]](#).

Responder Timestamp Format: The format of the timestamp values written by the responder, as specified in [Section 3.4 of \[RFC6374\]](#).

Responder's Preferred Timestamp Format: The timestamp format preferred by the responder, as specified in [Section 3.4 of \[RFC6374\]](#).

Session Identifier: Set arbitrarily in a query and copied in the response, if any. This field uniquely identifies a measurement operation (also called a session) that consists of a sequence of messages. All messages in the sequence have

same Session Identifier [[RFC6374](#)].

TC: Traffic Class being measured.

Timestamp 1-4 (T1-T4): Referring to Figure 2.

The mapping of timestamps to the Timestamp 1-4 fields is designed to ensure transmit timestamps are always written at the same fixed offset in the packet likewise for receive timestamps. This property is important for hardware processing.

TLV Block: Zero or more TLV fields. This document assumes the use of the DM message TLV defined in [[RFC6374](#)].

[3.2.4](#). One-Way Delay Measurement using Synthetic Probes

For delay measurement using synthetic probes, a DM TLV in the SRH to record timestamps and END.OTP SID as described in the pseudocode in [[I-D.draft-fils-spring-srv6-network-programming](#)] to punt the packet are used.

[3.2.4.1](#). Example Procedure

To measure one-way delay from node N1 over an SR Policy that goes through a segment-list (A2::C31, A4::C52) to node N4, following procedure is followed:

0 Node N1 constructs a DM probe packet with (B1::0, A2::C31)(A4::C52, A2::C SL=1; NH=NONE, DM TLV). To punt the DM probe packet at node N4, node N1 inserts the END.OTP SID [[I-D.draft-fils-spring-srv6-network-programming](#)] just before the target SID A4::C52 in the SRH. Thus, the packet as it leaves node N1 looks like (B1::0, A2::C31)(A4::C52, A4::OTP, A2::C31; SL=2; NH=NONE, DM TLV (with from N1)). The PM synthetic probe query message does not contain any payload data.

0 When node N4 receives the packet (B1::0, A4::OTP)(A4::C52, A4::OTP, A2::C SL=1; NH=NONE, DM TLV), it processes the END.OTP SID, as described in the pseudocode in [[I-D.draft-fils-spring-srv6-network-programming](#)]. In doing so, it punts the timestamped packet (with T2 from N4) to the Performance Measurement (PM) process for processing.

The PM process on node N4 responds to the DM probe message as following:

0 The Source Address object (Type=130) and Destination Address object (Type=

TLVs [[RFC6374](#)] indicate the addresses of the sender and the intended recipient of the PM message, respectively. The Source Address of a query message SHOULD be used as the destination unless an out-of-band response mechanism has been configured such as return controller's address is locally configured.

0 When a Return Address TLV object (Type=1) [[RFC6374](#)] is present in which the Return Address specifies the target address for the response message.

0 If the querier node N1 requires the response to be sent to the controller (N100), it adds the target controller's IP address in the Return Address TLV object of the DM message.

0 For one-way delay measurement, the querier node can send the UDP Return Object (URO) (Type=131) defined in [[RFC7867](#)]. From the responder node, the response is sent back to the PM querier node using the UDP Return Object (URO) TLV (Type=131) defined in [[RFC7867](#)] when the URO TLV is present in the PM query message. The process copies the content of the DM TLV into the payload of the PM reply message.

3.2.5. In-situ One-Way Segment-by-Segment Delay Measurement

For delay measurement for in-situ with data traffic, a DM TLV in the SRH to carry timestamps and O-bit as described in [I-D.[draft-filsfils-spring-srv6-network-programming](#)] to punt the packet on every SRv6 nodes are used.

3.2.5.1. Example Procedure

Consider the case where the user wants to measure one-way delay from node N1 to node N4 via an SR Policy that goes through a segment-list (A2::C31, A4::C52). However, the user desired to get the delay measurement done in-situ with data traffic on a segment-by-segment basis.

0 To force a punt of the time-stamped copy of the data packet at node N2 and N4, node N1 sets the O-bit in SRH at locally configured periodic measurement interval. The packet, as it leaves node 1, looks like (B1::0, A2::C31)(A4::C52, A2::C31; SL=1, Flags.O=1, DM TLV (with T1 from N1), NH=data payload type)(data payload). Here, the data payload refers to the actual data traffic going over the SR policy whose performance is being measured. Node N1 may optionally punt a time-stamped copy of the packet with T1 to the local PM process.

0 When node N2 receives the packet (B1::0, A2::C31)(A4::C52, A2::C31; SL=1, Flags.O=1, DM TLV, NH=data payload type)(data payload) packet, it processes the O-bit in SRH, as described in the pseudocode in [I-D.[draft-filsfils-spring-srv6-network-programming](#)]. A time-stamped copy of the packet gets punted to the local PM process for processing. Node N2 continues to apply the A2::C31 SID function to the original packet and forwards it, accordingly. As SRH.Flags.O=1, Node N2 disables the PSP flavour, i.e., does not remove the SRH.

0 The PM process at node N2 sends the copy of the time-stamped packet (with TLV containing T1 from N1 and T2 from N2) to a locally configured controller the querier. Please note that, as mentioned in [I-D.[draft-filsfils-spring-srv6-network-programming](#)], if node N2 does not support the 0-bit, it simply ignores and processes the local SID, A2::C31. In this case, the controller will not process the performance data from the segments with the nodes that do not support the 0-bit.

0 When node N4 receives the packet (B1::0, A4::C52)(A4::C52, A2::C31; SL=0, Flags.O=1, DM TLV (containing T1 from N1); NH=data payload type)(data payload type) processes the 0-bit in SRH, as described in the pseudocode in [I-D.[draft-filsfils-spring-srv6-network-programming](#)]. A time-stamped copy of the packet gets pushed to the PM process for processing.

0 The PM process at node N2 sends the copy of the time-stamped packet (with TLV containing T1 from N1 and T2 from N4) to a locally configured controller the querier.

The Controller processes the time-stamped packet from each segment and computes the segment-by-segment one-way delay. Support for 0-bit is part of node capability advertisement. That enables node N1 and the controller know which segment nodes are capable of sending time-stamped copy of the packet.

[4. Security Considerations](#)

TBA.

[5. IANA Considerations](#)

IANA is requested to allocate a value for the new SRH TLV Type for Delay Measurement.

[6. Contributors](#)

Faisal Iqbal
Cisco Systems, Inc.
Email: faiqbal@cisco.com

Carlos Pignataro
Cisco Systems, Inc.
Email: cpignata@cisco.com

[7. References](#)

[7.1. Normative References](#)

[RFC6374] Frost, D. and S. Bryant, "Packet Loss and Delay Measurement for Multicast Networks", DOI 10.17487/RFC6374, [RFC 6374](#), September 2011.

Internet-Draft

SR Performance Measurement

December 2017

[RFC7876] Bryant, S., Sivabalan, S., and Soni, S., "UDP Return Path for Packet Loss and Delay Measurement for MPLS Networks", [RFC 7876](#), July 2016.

[I.D-filsfils-spring-srv6-network-programming] SRv6 Network Programming, [draft-filsfils-spring-srv6-network-programming](#), C. Fisfils, work in progress.

[7.2](#). Informative References

[I-D.brockners-inband-oam-data] Data Formats for In-situ OAM. F. Brockners, work in progress.

[I-D.brockners-inband-oam-transport] Encapsulations for In-situ OAM Data, F.Brockners, work in progress.

[I-D.brockners-inband-oam-requirements] Requirements for In-situ OAM, F.Brockners, work in progress.

[8](#). Acknowledgments

To be added.

Internet-Draft

SR Performance Measurement

December 2017

Authors' Addresses

Clarence Filsfil
Cisco Systems, Inc.
Email: cfilsfil@cisco.com

Zafar Ali
Cisco Systems, Inc.
Email: zali@cisco.com

Rakesh Gandhi
Cisco Systems, Inc.
Email: rgandhi@cisco.com

Nagendra Kumar
Cisco Systems, Inc.
Email: naikumar@cisco.com

Dirk Steinberg
Steinberg Consulting
Germany
Email: dws@dirksteinberg.de

Stefano Salsano
Universita di Roma "Tor Vergata"
Italy
Email: stefano.salsano@uniroma2.it

ali, et al.

Expires June 23, 2018

[Page 12]