



Segment Routing Part II

SEGMENT ROUTING 详解

流量工程

(第二卷)

【比】克拉伦斯·菲尔斯菲尔斯 (Clarence Filisfilis)

【比】克里斯·米克尔森 (Kris Michielsen)

【法】弗朗索瓦·克拉德 (Francois Clad)

【加】丹尼尔·沃耶 (Daniel Voyer)

◎著

苏远超 钟庆◎译

蔡德忠 蒋治春◎审校



中国工信出版集团



人民邮电出版社
POSTS & TELECOM PRESS



Segment Routing Part II

SEGMENT ROUTING 详解

流量工程

(第二卷)

【比】克拉伦斯·菲尔斯菲尔斯 (Clarence Filisfilis)

【比】克里斯·米克尔森 (Kris Michielsen)

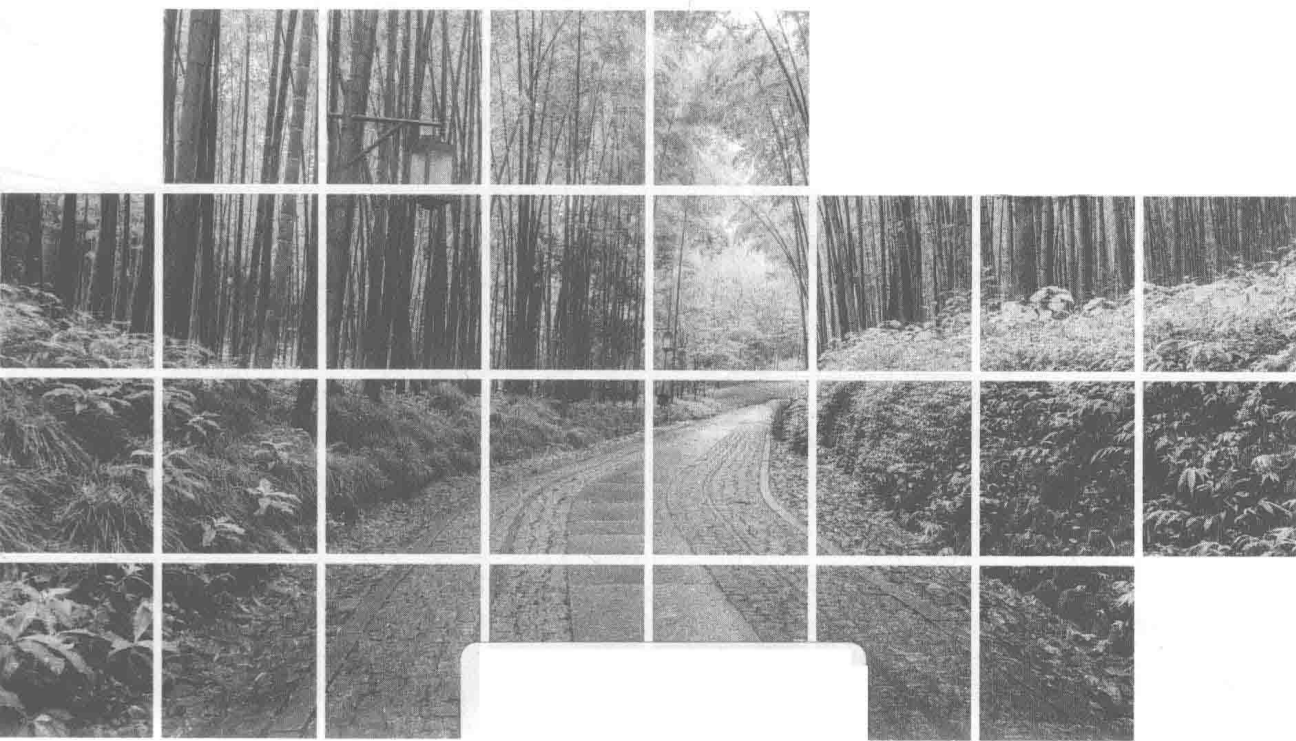
【法】弗朗索瓦·克拉德 (Francois Clad)

【加】丹尼尔·沃耶 (Daniel Voyer)

◎著

苏远超 钟庆◎译

蔡德忠 蒋治春◎审校



人民邮电出版社

北京

图书在版编目 (CIP) 数据

Segment Routing详解. 第二卷 / (比) 克拉伦斯·菲尔斯菲尔斯 (Clarence Filsfils) 等著; 苏远超, 钟庆译. — 北京: 人民邮电出版社, 2019.9
ISBN 978-7-115-51809-5

I. ①S… II. ①克… ②苏… ③钟… III. ①互联网
络—路由器—基本知识 IV. ①TN915.05

中国版本图书馆CIP数据核字(2019)第179728号

内 容 提 要

本书详细介绍SEGMENT ROUTING流量工程(以下简称为SR-TE)解决方案以及SDN控制器与SEGMENT ROUTING基础设施的集成,并提供了相关各种协议的教程。通过本书(本系列丛书第二卷),读者能够更加深入地理解SEGMENT ROUTING技术如何推动网络简化、扩展和自动化,如何结合SDN实现大规模流量工程,如何对业务进行监控和维护,提高网络可见性等。本书适合电信运营商、OTT、大中型企业、设备制造商、软件开发商、系统集成商等广大IP网络从业人员以及在校本科高年级学生、硕士、博士等阅读与学习。

版权声明

Segment Routing, Part II

Traffic Engineering by Clarence Filsfils, Kris Michielsen, Francois Clad and Daniel Voyer

Copyright © 2019 Cisco Systems, Inc. All rights reserved.

本书中文简体字版由 Cisco Systems, Inc. 公司授权人民邮电出版社出版, 专有版权属于人民邮电出版社。

- ◆ 著 【比】克拉伦斯·菲尔斯菲尔斯 (Clarence Filsfils)
【比】克里斯·米克尔森 (Kris Michielsen)
【法】弗朗索瓦·克拉德 (Francois Clad)
【加】丹尼尔·沃耶 (Daniel Voyer)

译 苏远超 钟庆

审 校 蔡德忠 蒋治春

责任编辑 赵娟

责任印制 彭志环

- ◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号
邮编 100164 电子邮件 315@ptpress.com.cn
网址 <http://www.ptpress.com.cn>
北京市艺辉印刷有限公司印刷

- ◆ 开本: 800×1000 1/16

印张: 34

2019年9月第1版

字数: 736千字

2019年9月北京第1次印刷

著作权合同登记号 图字: 01-2019-5312号

定价: 168.00元

读者服务热线: (010)81055488 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京东工商广登字20170147号

精彩推荐

“随愿网络 (IBN)” 是继 SDN/NFV 后网络发展的下一目标。“随愿网络” 的落地要求网络基础设施具备高度的扩展性、灵活性和敏捷性。过去数年的快速发展表明, Segment Routing 将是支撑“随愿网络” 的基本构建模块。《Segment Routing 详解 (第二卷)》聚焦于 Segment Routing 流量工程的创新及应用, 这对于正在进行的网络转型改造、5G 承载 / 切片、云网协同都有着重要的指导作用和参考价值。本书译者和审校者均为国内 Segment Routing 技术的先行者和实践者, 对 Segment Routing 流量工程有着深入的理解, 翻译准确细致, 值得向广大读者推荐。

——SDN/NFV/AI标准与产业推进委员会主席, 工业和信息化部科技委常务副主任,
中国电信集团公司科学技术委员会主任 韦乐平

Segment Routing 在过去几年得到业界的迅速认可与发展, 体现了新技术强大的生命力。网络的流量工程 TE 技术一直是学术界和工业界共同关注的尖端课题。20 多年前基于 MPLS 架构所提出的 RSVP-TE 技术后来证明在工业实践中依然面临一系列挑战, 并没有得到大规模的采用, 原因在于其技术过于复杂, 可扩展性比较差。近几年, Segment Routing 作为 MPLS 架构新一代的更新技术, 学术界和工业界都非常关注其在网络流量工程 TE 方面的技术原理和实用改善。最近, 国内和国外领先企业也都开展了 SR-TE 的商用实践并取得了明显的成果。《Segment Routing 详解 (第二卷)》专门研究和系统阐述了 Segment Routing 流量工程 SR-TE, 这对于深入理解 Segment Routing 技术架构及其在流量工程 TE 领域的实践大有裨益。本书原版继续由 Segment Routing 发明者 Clarence 亲自参与撰写, 由广受好评的第一卷的译者 / 审校者继续担纲翻译、审校, 是学习 Segment Routing 和 SR-TE 的必读书籍, 特诚意向各位读者推荐!

——清华大学电子工程系教授、博士生导师,
中国教育和科研计算机网 (CERNET) 国家网络中心副主任 李星

思科公司自 2013 年提出 Segment Routing 技术以来, 不断引领 Segment Routing 技术的发展, 并致力于推动技术标准化、达成行业共识以及实际应用部署。可以说, 在众多

的应用场景中, Segment Routing 已经成为事实上的 SDN 标准架构, 是新一代“基于意图的网络”的核心组成部分。作为 Segment Routing 最主要的创新与价值体现, Segment Routing 流量工程已经在全球众多客户的网络中得到了实际部署, 并且部署的速度在加快, 部署的规模在扩大, 方兴未艾, 前景光明。为此, 业界系统介绍 Segment Routing 流量工程的权威书籍《Segment Routing 详解 (第二卷)》的出版恰逢其时。本书是由 Segment Routing 发明人兼思科院士 Clarence 等人撰写, 由浅入深、内容详实、实用性强, 是深入了解 Segment Routing 的必读及参考书籍。思科资深专家苏远超、钟庆第一时间将其翻译为中文并呈献给国内业界, 以飨读者。本书适合网络相关从业人员、大中专院校相关专业师生阅读。

——思科全球副总裁, 大中华区首席技术官 曹图强

运营商正在进行网络重构, 以支撑业务可持续发展的战略。为应对多云、云原生、5G、物联网所带来的超大规模、智能、端到端流量调度需求, 网络传送技术需要进行根本性的创新, Segment Routing 无疑是业界在这方面的共同选择。《Segment Routing 详解 (第二卷)》完整展现了 Segment Routing 流量工程技术的全貌, 既有技术设计思想介绍, 又有协议实现细节和部署建议, 具有很高的参考价值。第二卷在第一卷的基础上进一步深化与现实结合的技术实施场景, 具有很高的学术和工程参考价值。本书译者具有多年运营商网络经验, 对 Segment Routing 技术有着非常深入的研究, 翻译严谨细致, 使本书通俗易懂, 适合广大电信网络从业人员阅读。

——中国电信北京研究院院长 陈运清

中国联通在国内率先基于 Segment Routing 技术推出了“云联网”产品, 实现了多云环境下的智能云网协同, 成为产业互联网的标杆网络, 实现了良好的经济效益和社会效益。中国联通正在基于 Segment Routing 构建下一代的 5G 承载网和城域网, 因此《Segment Routing 详解 (第二卷)》所介绍的流量工程内容非常具有实用性; 里面介绍的灵活算法功能对于实现 5G 切片亦有重要的借鉴意义。第二卷原版继续由 SR 发明人 Clarence 联合其他专家撰写, 并继续由思科资深专家苏远超等翻译, 具有很高的参考价值。

——中国联通网络技术研究院首席专家 唐雄燕

随着云和行业应用的发展, 网络需要为海量用户和应用提供差异化服务, 并为业务提供强有力的支撑。Segment Routing 能够与 SDN 完美结合, 把网络打造成智能化平台。尤其是 Segment Routing 流量工程, 更是提高用户体验不可或缺的强有力工具。欣

闻《Segment Routing 详解（第二卷）》出版，这是市面上 Segment Routing 流量工程的权威专著，是广大网络从业者必备的指导书、参考书和工具书。本书继续由第一卷相关专家进行翻译、审校，延续了第一卷严谨细致的作风，翻译准确到位，特向各位读者推荐。

——腾讯网络平台部网络架构中心总监，专家工程师 邵华

在《Segment Routing 详解（第一卷）》推出后不久，很高兴看到思科专家又给我们带来了第二卷。本书在第一卷的基础上，结合实际应用场景，详细阐述了基于 SR 的流量工程（SR-TE）的一系列关键技术：多约束条件下路径计算、SR Policy、引流等。作者凭借其深厚的理论功底和丰富的实践经验，在提出一个想法或解决方案时，从问题的本质入手，抽丝剥茧，引经据典，层层推进，娓娓道来，增添了本书阅读的趣味性。中文译者保持了一贯的高水准，翻译精准严谨。随着 Segment Routing 在广域网、城域网，甚至是数据中心越来越广泛的应用和部署，本书的出版可谓恰逢其时，不仅能帮助读者掌握其核心技术，而且有助于推动网络的持续创新和演进。

——大地云网联合创始人、CTO 鲁子弈

译者序

《Segment Routing 详解 (第一卷) 》受到了广大读者的欢迎, 成为学习 Segment Routing 必备的工具书和参考书, 非常感谢读者们的支持!

距离《Segment Routing 详解 (第一卷) 》出版已经过去两年时间, 在此期间, Segment Routing 标准化得到了迅速提高, 很多厂商的互操作取得了长足进步, 国内外关于 Segment Routing 部署的案例增长迅速。尤其是, 中国成为全球 Segment Routing 研究和部署最热的国家之一。对此, 我们感到非常欣慰, 也坚定了我们继续翻译后续第二卷、第三卷的信心和使命感。

本书是《Segment Routing 详解 (第二卷) 》, 聚焦于 Segment Routing 流量工程 (SR-TE)。SRv6 相关内容将会在第三卷中介绍。

SR-TE 无疑是 Segment Routing 最重要的应用之一, 也是很多用户部署 Segment Routing 的主要动机之一 (另一个是 TI-LFA 快速收敛)。众所周知, 流量工程并不是新鲜事物, 我们在 2004 年设计建设中国电信 CN2 时, MPLS 流量工程就是重点关注的内容, 我们最终在 CN2 上部署了基于 RSVP-TE 的 MPLS FRR。国外 RSVP-TE 的部署案例更是不少。但随之而来的 N^2 数量级的隧道及状态使网络的稳定性急剧降低, 对很多使用 RSVP-TE 的用户造成了十足的困扰。

因此, SR-TE 源路由及只在网络边缘维持状态的特点迅速被业界所接受, 并在运营商、OTT、企业网络中得到大量部署。但需要指出的是, 目前部署的绝大多数 SR-TE 都是隧道接口模式 (Tunnel Interface), 基本上沿袭了传统 RSVP-TE 的做法, 业务标识、引流、切片非常麻烦; 在路径计算算法方面, 仍然有基于电路而非 IP 的实现存在, 这不但使网络中的等价路径不能得到充分利用, 还导致 Segment 列表长度增加, 对现网设备要求过高。

为了满足 5G、万物互联、多云对网络提出的超大规模、超高可用性、超高传送效率和超灵活的要求, 实现向基于意图的网络 (IBN) 转型, 思科在 2017 年提出新的 SR-TE 模型, 即 SR Policy, 其目前已经成为 IETF 工作组草案; 业界其他厂商也在积极跟进并支持 SR Policy。SR Policy 并非传统隧道接口模式流量工程的简单升级, 而是从头重新设计的一套完整体系: SR-TE 不再对应于隧道接口, 而对应于表示意图的 Segment 列表; BGP 路由触发、

自动引流、按需下一跳 (ODN)、灵活算法 (Flex-Algo)、不相交路径 (Disjoint-path) 等激动人心的强大功能都是 SR Policy 的基本组件, 并且各个组件可以像乐高积木一样组合起来使用; SR Policy 通过 PCEP/BGP SR-TE 与控制器结合, 采用 Segment Routing 原生算法计算路径, 完美地实现了分布式智能和集中式优化的结合, 从而将 SR-TE 的威力发挥到极致, 使构建最优的端到端跨域流量工程路径变得简单易行; 不断增强的性能测量和弹性则保证了 SR Policy 可用于承载关键业务。

本书的内容非常新, 且贴近工作实际。我们在一年前就开始与原著作者进行大量的沟通, 基于我们在国内的实际部署经验及客户的反馈, 在内容、章节组织、功能描述等方面提出建议。希望本书能更好地适用于国内的部署要求, 更具可操作性。然而由于技术水平有限, 时间匆忙, 纰漏之处在所难免, 还望各位读者不吝指正。

本书的出版要感谢思科众同事的支持: 曹图强 (Tuqiang Cao)、方芳 (Flora Fang)、Kevin Wollenweber、孙强 (Victor Sun)、李巍 (Vivian Li) 等, 这本书的出版离不开他们的帮助。同时也感谢阿里巴巴蔡德忠 (Dennis Cai) 先生和腾讯蒋治春 (Roy Jiang) 先生的认真审阅, 为本书的质量把关。还需要感谢我们的家人, 他们对于我们用业余时间翻译这本书给予了充分的理解和支持。最后必须感谢人民邮电出版社专业的编辑团队, 给了我们很多建议, 令本书增色不少。

苏远超 钟庆
2019年6月于广州

前言

本书是《Segment Routing 详解》系列丛书的第二卷。

随着越来越多的运营商和企业 在单一的网络基础设施上支持日益增多的业务类型，定制适合应用传送需求的能力变得至关重要。

在这方面，操作员多年以来一直在探索流量工程技术，但是明显遇到了很多可扩展性问题，使他们无法对所提供的大量业务实现端到端、细颗粒度的控制。

Segment Routing Traffic Engineering (SR-TE, SR 流量工程) 彻底改变了这一状况，并已经成为提供大规模流量工程能力无可争议的解决方案。

面向的读者

我们试图让本书可面向广泛的读者群，书中会讨论初级、中级和高级主题。我们希望任何尝试从实际应用角度设计、支持或者只是理解 SR-TE 的读者，都能从本书中获益。这包括运营商和企业的网络设计人员、工程师、管理员和操作员，以及其他希望了解 SR-TE 的专业人士或者学生。

我们假设读者熟悉数据网络，了解 IP、IP 路由以及 MPLS 的概念，关于这些方面的内容有很多好书可供参考。

我们还假设读者了解 Segment Routing 的基础知识。本书第一卷(amazon.com 有售) 是很好的学习 SR 基础知识的资源，中文版已由人民邮电出版社于 2017 年 8 月出版。

免责声明

本书仅反映作者的观点，与作者所属公司无关。本书中每项陈述都是从作者个人研究和

实验室测试得出的结论。

还要注意以下事项。

- 一些示例采用原型软件版本构建。在本书出版之时，一些功能和命令可能尚未正式支持。思科公司不承诺发布本书中介绍的任何功能。对于某些功能，文中有明确说明，但并不是所有功能都有。从好的方面来看，本书提供了对最新技术的预览，从而使读者有机会体验可能会出现的技术。
- 本书中使用的某些命令在将来可能会被更改或者废弃。本书不保证语法的准确性。
- 为了便于说明，作者自行编辑了一些命令输出示例，例如删除部分文字。因此，本书不保证示例的准确性。

审稿人

许多人为本书做出了贡献，在第1章中将向他们致谢。为了写出一本准确、清晰、好看的书，除了作者本身需要仔细校对之外，还需要很多人的帮助。

在此，我们要特别感谢那些在本书各个阶段帮忙审稿的人们。没有他们，这本书就不能达到目前的水准。真诚地对所有人说一声：“谢谢！”

审稿人 (以姓氏字母为序)：

Gaurav Dawra, Mike DiVincenzo, Alberto Donzelli, Darren Dukes, Muhammad Durrani, Rakesh Gandhi, Arkadiy Gulko, Al Kiramoto, Przemyslaw Krol, Sakthi Malli, Paul Mattes, Hardik Patel, Rob Piasecki, Carlos Pignataro, Robert Raszuk, Joel E. Roberts, JC Rode, Aisha Sanes, Grant Socal, Simon Spraggs, 苏远超 (YuanChao Su), Ketan Talaulikar, Mike Valentine, Bhupendra Yadav, 赵勇 (Frank Zhao) 和钟庆 (Qing Zhong)。

文本约定

本书内容组织包含多种样式。

- 一般样式：这是学习 Segment Routing 的读者遵循的常规流程。它包含的是事实，是客观内容，而不是观点。
- 重点提示：重点提示框用于强调重要的元素和主题。这些显示在“重点提示”框中。



重点提示

这是一个重点提示的例子。

- 观点：观点框用于表达观点、选择和折衷。此内容不是理解 Segment Routing 所必需的，但为感兴趣的读者提供了更多的背景信息，并以引号形式呈现。我们还邀请了在 Segment Routing 项目上非常积极的业内同行，就 Segment Routing 总体情况或者一些具体方面发表意见。这些引言后面附有提供此观点者的姓名。

“这是一个观点的例子”。

— John Doe

- 提示：提示框用于简要说明有助于理解一般样式的技术信息（大部分与 Segment Routing 本身无关）。这些内容显示在“提示”框中。

提示

这是一个提示的例子。

插图和示例约定

本书中的插图和示例遵循以下约定。

- 节点 X 的路由器 ID 为 1.1.1.X。其他环回地址为 1.1.n.X，其中 n 为索引。
- 节点 X 连接到节点 Y 的接口的 IPv4 地址为 99.X.Y.X/24，其中 $X < Y$ 。例如节点 2 连接到节点 3 的链路具有网络地址 99.2.3.0/24；节点 2 上的接口地址为 99.2.3.2，节点 3 上的则为 99.2.3.3。
- Prefix-SID 是在 [16000-23999] 范围内的标签。这是思科设备的默认 SRGB。
- 显式本地 Segment 是在 [15000-15999] 范围内的标签。这是思科设备的默认 SRLB。

- 动态 Adj-SID 是在 [24000-24999] 范围内的标签, 节点 X 上用于到节点 Y 的邻接的 Adj-SID 格式为 240XY。
- 动态 BSID 是在 [40000-49999] 范围内的标签。
- 动态 BGP EPE Peering-SID 是在 [50000-50999] 范围内的标签。
- 其他 MPLS 应用 (例如 LDP、RSVP-TE、BGP-LU 等) 分配的动态标签在 [90000-99999] 的范围内。

目 录

第 1 章 简介

1.1 第一卷的主观性简介	1
1.2 术语	1
1.3 设计目标	2
1.3.1 IP 优化的解决方案	2
1.3.2 简单的解决方案	2
1.3.3 可扩展的解决方案	3
1.3.4 模块化的解决方案	3
1.3.5 创新的解决方案	5
1.4 服务等级保证	5
1.5 流量矩阵	6
1.6 容量规划	7
1.7 对集中式计算的依赖	7
1.7.1 不相交路径	8
1.7.2 跨域	8
1.7.3 带宽代理	8
1.7.4 多层优化	9
1.8 用 Segment/Segment 列表表达流量工程意图	9
1.9 SR Policy	11
1.10 BSID	11
1.11 需要多少个 Segment	12
1.12 基于着色业务路由的自动化	13
1.13 SR-TE 进程	14
1.13.1 一个进程，多重角色	14

1.13.2	组件	15
1.13.3	SR-TE 数据库	16
1.13.4	SR 原生算法	16
1.13.5	与其他进程和外部 API 的交互	17
1.13.6	新的命令行	18
1.14	服务编程	18
1.15	领先运营商团队	19
1.16	SR-TE 思科团队	20
1.17	标准化	21
1.18	本书内容组织顺序	22
1.19	参考文献	22

第 2 章 SR Policy

2.1	简介	25
2.1.1	SR Policy 的显式候选路径	27
2.1.2	路径验证和选择	29
2.1.3	低延迟动态候选路径	31
2.1.4	避免特定链路的动态候选路径	32
2.1.5	在 Segment 列表中编码路径	34
2.2	SR Policy 模型	35
2.2.1	Segment 列表	36
2.2.2	候选路径	36
2.3	BSID	39
2.4	SR Policy 配置	41
2.5	小结	46
2.6	参考文献	47

第 3 章 显式候选路径

3.1	简介	49
3.2	SR-MPLS 标签	50
3.3	Segment 描述符	53
3.4	验证显式候选路径	56
3.5	实际应用中的考虑	60

3.6	控制器发起的显式候选路径	61
3.7	TDM 迁移	62
3.8	使用 Anycast-SID 的双平面不相交路径	65
3.9	小结	70
3.10	参考文献	70

第 4 章 动态候选路径

4.1	简介	71
4.1.1	表示动态路径的目标和约束条件	72
4.1.2	计算路径 = 解决最优化问题	72
4.1.3	SR 原生算法与基于电路算法的对比	73
4.2	分布式计算	74
4.2.1	头端计算低延迟路径	75
4.2.2	头端计算带约束条件路径	79
4.2.3	其他的应用场景和限制	83
4.3	集中式计算	84
4.3.1	SR PCE	84
4.3.2	SR PCE 计算不相交路径	86
4.3.3	SR PCE 计算跨域端到端路径	93
4.4	小结	101
4.5	参考文献	102

第 5 章 自动引流

5.1	简介	104
5.2	BGP 路由着色	106
5.2.1	BGP 颜色扩展团体属性	106
5.2.2	在出口 PE 处对 BGP 路由着色	107
5.2.3	颜色使用冲突	109
5.3	VPN 前缀自动引流	110
5.4	引导具有不同 SLA 的多个前缀	118
5.5	EVPN 自动引流	123
5.6	其他业务路由	127
5.7	禁用自动引流	127

5.8 适用性	128
5.9 小结	129
5.10 参考文献	130

第 6 章 按需下一跳

6.1 着色	132
6.2 按需生成候选路径	132
6.3 与 SR-TE 解决方案无缝集成	133
6.4 拆除 ODN 候选路径	134
6.5 示例说明: 域内 ODN	134
6.6 示例说明: 跨域 ODN	139
6.7 ODN 仅用于授权颜色	142
6.8 小结	143
6.9 参考文献	144

第 7 章 灵活算法

7.1 Prefix-SID 算法	146
7.2 算法定义	152
7.2.1 一致性	154
7.2.2 定义通告	155
7.3 路径计算	159
7.4 TI-LFA 备份路径	161
7.5 与 SR-TE 集成	164
7.5.1 ODN/ 自动引流	166
7.5.2 跨域路径	168
7.6 双平面不相交路径应用场景	169
7.7 Flex-Algo Anycast-SID 应用场景	176
7.8 小结	182
7.9 参考文献	183

第 8 章 网络弹性

8.1 本地故障检测	186
------------------	-----

8.2	域内 IGP 泛洪	189
8.3	跨域 BGP-LS 更新	191
8.4	显式路径的验证	193
8.4.1	Segment 表示为 Segment 描述符	194
8.4.2	Segment 表示为 MPLS 标签	197
8.5	头端重新计算动态路径	199
8.6	SR PCE 重新计算动态路径	202
8.7	成员 Prefix-SID 的 IGP 收敛	203
8.7.1	IGP 回顾	203
8.7.2	显式候选路径	203
8.7.3	动态候选路径	205
8.8	利用 Anycast-SID	207
8.9	TI-LFA 保护	209
8.9.1	成员 Prefix-SID	210
8.9.2	成员 Adj-SID	212
8.9.3	TI-LFA 适用于 Flex-Algo Segment	215
8.10	不受保护的 SR Policy	215
8.11	其他机制	217
8.11.1	SR IGP 微环路避免	217
8.11.2	SR Policy 存活性检测	219
8.11.3	用于 SR Policy 中间 Segment 的 TI-LFA 保护	219
8.12	同时作用	220
8.13	小结	221
8.14	参考文献	222

第 9 章 BSID 和 SRLB

9.1	定义	224
9.2	显式分配	228
9.3	简化 / 可扩展	234
9.4	网络隐藏 / 业务独立	237
9.5	引导至远端 RSVP-TE 隧道	238
9.6	小结	239
9.7	参考文献	240