

MPLS+SDN+NFV World @ 巴黎 2018——Cisco IOS XR 参与互操作测试

作者：思科首席技术市场工程师 Jose Liste

又一次互操作测试

在所有 EANTC 主持的互操作测试中，我已经代表思科参与了超过 8 年。自 2008 年起，我已经在柏林度过了多个愉快的冬季，数量远远超出了我的想象。

仔细地回想起来，我见证了若干重塑网络产业的技术（甚至政治）斗争和演变，非常令人慨叹。很多技术、乃至整个公司在短时间内失败。在我看来，21 世纪第一个十年后期网络产业的关注点严重发散，劳而无功。还记得 PBT、PBB-TE、T-MPLS、MPLS-TP 及 OpenFlow 吗？

然而，有一项技术却一直保持着旺盛的生命力——那就是 MPLS。在 OpenFlow 大肆炒作时期，几个所谓的专家质疑运营商网络中 MPLS 的未来。那时，思科正在开启 Segment Routing (SR) 的旅程。2012 年 11 月，思科院士 Clarence Filisfilis 首次向网络运营商们披露了 SR 的概念。仅仅一年之后，IETF 就正式成立了一个新的网络工作组——源数据包路由网络 (SPRING)。2015 年，EANTC 开展了第一个公开的 SR 互操作测试活动。时至今日，这个互操作测试证明了众多厂商已经提供对 SR 技术的强有力支持。

另一个展示了成功轨迹的技术则是以太网 VPN (EVPN)。EVPN 最初定位为 L2VPN 的下一代解决方案，尤其针对二层数据中心互联 (DCI) 以及跨多点以太网 LAN (E-LAN) 业务。后来，EVPN 快速地演进，以适应其它应用场景的需求，例如点到点以太网专线 (E-LINE) 和点到多点以太网 (E-Tree) 业务。目前，EVPN 的应用案例已经不止是运营商级以太网，其已经深入应用到数据中心内部，用于子网内和子网间转发。2013 年，我参与了 EANTC 第一次公开的 PBB-EVPN 互操作测试。5 年以来，多厂商支持以及测试场景在不断成熟。

历史就谈到这里，下面让我们转到这篇博客的重点部分。

本文旨在提供思科参与今年互操作测试的技术概述，主要针对使用 IOS XR 操作系统的平台。技术方面，我重点关注 SR 和 EVPN 相关测试。需要注意的是，思科数据中心产品线以及网络服务编排器 (NSO) 也参与了测试。

首先，我强烈建议读者阅读 [EANTC 的官方公开白皮书](#)。本博客中，除了补充白皮书的内容，我还希望能够后退一步，针对测试结果提供更宽广的背景和视角——**为什么我应该关注这个测试活动？这些测试结果代表什么真实情况？**最后，我会提供进一步的见解，运营商在评估厂商产品时应该注意什么——**除了报告的结果外还应该关注什么？**

总体情况

今年的测试，有史以来参与厂商数量最多（总共 21 个），参测设备超过 60 种，有望成为对网络运营商意义重大的活动。在我看来，它也的确达成了这个预期。



MPLS WC 的 Interop 展台-照片由 EANTC 提供©

下面列出了此次测试中思科 SR 的相关信息：

- 思科是参与验证 SR 实现就绪情况的十家网络和测试设备厂商之一。所有的主要网络厂商均参与了互操作测试
- 目前，SR-MPLS 在使用 MPLS 作为传送的测试例中处于主导地位，包括 IP VPN 业务和 EVPN 业务的传输。测试中尽可能少地使用 LDP，而 RSVP-TE 则完全未使用
- 整个测试中，IS-IS 被选为主要的 IGP 协议。需要注意的是，测试曾考虑使用 OSPF，但是由于时间限制，没有被优先考虑。

- **测试成功验证了基本的 IS-IS SR 扩展**，思科及各厂商之间没有发现互操作问题。验证的功能包括：
 - IPv4 控制平面
 - 主机前缀的 Prefix Segment ID (Prefix-SID)，包括 Node SID 和 Anycast SID
 - IS-IS 邻接的 Adjacency Segment ID (Adj-SID)
 - SR 映射服务器 (SRMS) 功能，执行前缀与 SID 映射关系的通告
- **SR 流量工程 (SRTE) 是另一个关注领域**，验证了以下内容：
 - 路径计算单元协议 (PCEP) ——有状态 PCE 模型
 - 针对 Segment Routing 的 PCEP 扩展
 - 针对 Segment Routing 的 BGP 链路状态协议 (BGP-LS) 以及扩展
- 此外，**EANTC 第一次验证了以下 SR-MPLS 相关内容**：
 - 与拓扑无关的 LFA (TI-LFA)
 - **BGP SR Prefix SID 扩展 (SR BGP)**
 - **SR 运营、管理以及维护 (OAM)**
- 最后，EANTC 也首次验证了 **SRv6**。测试涵盖了 IETF 草案 “[SRv6 Network Programming](#)” 中的基础功能

本博客稍后介绍 EVPN 相关情况。

思科参测设备

今年的活动中，思科 **ASR9000** 和 **NCS5500** 产品家族代表 Cisco IOS XR 产品系列参测。**NCS5500** 是首次亮相，**Cisco IOS XRv9000** 虚拟路由器也是首次作为虚拟 SR 路径计算单元(PCE)参与。

下面我们探讨主要的测试分类。

基于 Segment Routing 的 TI-LFA

对运营商而言，提供有 SLA 保证的业务，而且能够在网络组件故障情况下自动恢复连接是至关重要的。与拓扑无关的 LFA (TI-LFA) 依靠 Segment Routing，提供了一种本地修

复机制来实现这个目标。通过 IETF 草案规定的行为，TI-LFA 提供了一些关键优势，具体包括：

- **基于目的地址的自动保护**——IGP 针对每个目的地址（前缀）预先自动计算备份路径。连接到目的地的链路出问题，TI-LFA 激活数据平面的切换
- **TI-LFA 涵盖范围**——TI-LFA 为任意拓扑提供小于 50 毫秒的链路、节点以及本地 SRLG 保护。无论故障发生前后拓扑如何，TI-LFA 均能提供无环备份路径
- **最优路由**——TI-LFA 要求备份路径与路由收敛后路径一致，从而提供了最优路由
- **无状态操作**——基于源路由机制，实施备份路径不用在网络中创建额外的转发状态

由于这些优势，同时基于我们的部署经验，目前 TI-LFA 仍然是 SR 部署的主要驱动力之一。自从 2014 年思科支持 TI-LFA 以来，它一直是思科关键执行的技术领域之一。

因此，我们欢迎 EANTC 第一次在互操作测试中加入 TI-LFA 测试用例，思科的亮点包括：

- **思科成功验证了基于 TI-LFA 的小于 50 毫秒保护**
- **思科成功验证了 TI-LFA 链路保护**
- **思科成功验证了 TI-LFA 的本地 SRLG 保护。思科是唯一通过这个测试例的参测厂商**
- 需要注意的是，思科的实现也支持节点保护 TI-LFA，但这不属于测试计划的一部分。请继续关注思科即将在 2018 年夏季发布的 TI-LFA 增强功能！

最后，以下是一些在报告中**未能涵盖**，但在评估 TI-LFA 时却始终必须考虑的关键方面：

- 厂商实现是否提供针对每个目的地址（前缀）计算的备份路径？留意那些可能会投机取巧的地方，例如不针对每个目的地址计算最佳备份路径的实现。思科 TI-LFA 实现的设计满足这一目标
- 厂商实现是否提供与前缀无关的收敛？确保在激活备份链路期间，验证性能不随所保护前缀数量的增加而降低。思科 TI-LFA 实现遵循这个理念设计和实施
- 厂商实现是否提供对原本通过其它技术（如 LDP 信令或者纯 IP 路由）转发的流量的保护？确保验证涵盖了非 SR 的流量。思科 TI-LFA 实现能够用来保护 LDP 和 IP 流量

欲了解更多的信息，建议参考 [TI-LFA 教程](#) 以及 [TI-LFA 演示](#)。

SRTE 和 PCEP

Segment Routing 解决流量工程问题的能力是该技术最受欢迎的应用之一。

在 IETF 草案 “SRTE architecture” 中，详细说明了头端节点使用“SR 策略”，引导流量经过指定网络路径的行为和机制。基于源路由机制，所有的状态以有序 Segment 列表的方式编码在头端节点。因此，SRTE 与传统的流量工程解决方案不一样，不需要在中间节点维护状态。

SR 策略中的路径可以有不同的来源，其中包括由集中式控制器 PCE 进行计算。有一个 IETF 草案描述了 **PCEP 的 SR 扩展**，允许有状态 PCE 计算和发起流量工程路径，也允许路由器（PCC）向 PCE 发起请求，请求一条 SR 网络中满足特定限制条件和优化目标的路径。

总的来说，这是互操作测试中最活跃的一个领域，也是我个人花费时间最多的部分。EANTC 报告了超过 30 个成功的不同 PCE-PCC 厂商/产品组合，是获得最多正面结果的测试之一。

思科参与该测试领域的产品分为 2 类：作为 PCC 和作为 PCE。

思科作为 PCC 参与的亮点包括：

- 思科是 6 家参与 PCEP PCC 头端节点测试的厂商之一（不包括流量模拟器厂商）
- 在测试中，思科的 SR PCC 实现是互操作性最好的（考虑对接参测 PCE 厂商的测试成功数量）
- 作为 PCC，基于 PCE 发起模型，思科成功验证了与所有参测的非思科 SR PCE 对接以及创建、更新和删除 SR 策略
- 作为 PCC，基于 PCC 发起模型，思科成功验证了与所有参测的非思科 SR PCE 对接以及创建、更新和删除 SR 策略

思科作为 PCE 参与的亮点包括：

- 思科是 3 家参与 PCEP PCE 节点测试的厂商之一（不包括流量模拟器厂商）
- 思科的 SR PCE 在测试中是互操作性最好的（考虑对接参测 PCC 厂商的测试成功数量）
- 作为 PCE，基于 PCE 发起模型，思科成功验证了与非思科 SR PCC 对接以及创建、更新和删除 SR 策略
- 作为 PCE，基于 PCC 发起模型，思科成功验证了与非思科 SR PCC 对接以及创建、更新和删除 SR 策略
- 此外，思科 SR PCE 还成功验证了利用非思科节点产生的 BGP-LS 通告信息，学习到单一网络域和多网络域的拓扑
- 最后，测试中思科 SR PCE 唯一成功验证了在多域网络下的路径计算，其中在网络域边界利用到了出口对等体工程（EPE）SID



与来自华为（左）和诺基亚（右）的同事在 MPLS WC 演讲 © 照片由 EANTC 提供

最后，除了协议的互操作以外，运营商在评估 SRTE 头端节点和 PCE 实现时，考虑下面这些在报告中**未能涵盖**的关键方面也很重要：

- PCE 实现的路径计算是否符合 SR 原则——即是否最大限度地使用可用的 ECMP 路径和最小化标签栈大小？需再次留意那些可能投机取巧的地方，例如简单地重用 RSVP-TE 算法来实现 SR。RSVP-TE 是基于电路的，不支持 ECMP，因此在编码一条 SR 路径时需要很多的 SID。思科开发了用于 SR 路径计算的新算法，并获得了[学术界的认可](#)
- 厂商实现是否允许在头端节点进行路径计算？在大多数单域场景下，头端节点可用于计算路径。因此运营商应该不仅在 PCE，同时也可在 PCC 进行路径计算。请记住，头端节点和 PCE 节点的主要区别仅在于拓扑数据库的范围和大小。前者是单域网络，而后者可能是多域网络。思科的 SRTE 头端节点与 PCE 采用相同的计算算法
- 该实现是否提供了最大的扩展性，不需要预先建立全网状连接？思科的 SR 按需下一跳（SR ODN）技术在头端节点提供了一个本地机制，当被业务路由(例如 IP VPN)触发时，ODN 能够触发头端节点建立一条满足业务传送 SLA 要求的 SR 策略。[这里可了解更多信息及观看 ODN 演示](#)
- 该实现是否避免了复杂且极大影响性能流量引导技术？基于 IEFT 草案“SRTE architecture”规定的[引导行为](#)，思科支持一种创新的流量引导技术，我们称之为自动引导（Auto Steering）。自动引导基于业务路由的颜色（color）就可自动地把业

务流量引导到正确的 SR 策略。这个解决方案非常简单易用，且完全不影响性能。自动引导适用于采用不同方法创建的所有 SR 策略（例如按需的、PCE 发起的、本地配置的）。在上一点提到的 ODN 演示中也可观看自动引导演示。

欲了解更多信息，建议查看此 [SRTE 教程](#)

SR 和 LDP 互操作

SR 解决的一个关键应用场景是支持在现有网络中部署。IETF 草案“[Segment Routing interworking with LDP](#)”描述了几种在 SR 和 LDP 路由器混合组网环境下的 SR 和 LDP 互操作机制。

在 SR 和 LDP 位于网络不同部分的场景中，从 SR 到 LDP 方向，采用所谓 **SR 映射服务器 (SRMS)** 的功能生成端到端的 MPLS LSP。SRMS 是一个 IGP 节点，它可以通告其它 IGP 节点的前缀和 SID 的映射。

思科的 SR 实现从 2014 年开始就支持 SRMS 和 SR/LDP 在数据层面的互通。

思科参与此测试例的亮点包括：

- 思科成功验证了当作为只启用 SR 的节点时，接收到来自非思科 SRMS 的 IS-IS SRMS 通告
- 思科成功验证了在由非思科、只启用 SR 的节点所组成的网络中充当 SRMS 节点
- 思科成功验证了充当 LDP/SR 互操作的“拼接”节点

欲了解更多信息，建议参考此 [SRMS 教程](#)，以及 [SR/LDP 互通教程](#)。

BGP 的 SR 前缀 SID 扩展 (SR BGP)

IETF 草案“[Segment Routing Prefix SID extensions for BGP](#)”定义了一个 BGP 属性，用以通告 BGP Prefix SID 信息。BGP Prefix-SID 是全局 Segment，这个标识代表着要经由多条负载均衡的 BGP 最佳路径转发至对应前缀的指令。

IETF 草案“[BGP-Prefix Segment in large-scale data centers](#)”和“[Interconnecting Millions Of Endpoints With Segment Routing](#)”说明了 BGP Prefix SID 的应用案例。

该测试例是 EANTC 第一次进行 SR BGP 互操作测试。思科参与此测试例的亮点包括：

- 思科成功验证了在多厂商 SR BGP 交换矩阵中充当叶子节点(Leaf)
- 思科成功验证了在多厂商 SR BGP 交换矩阵中充当主干节点(Spine)

SR 操作、管理和维护 (OAM)

网络运营商需要具备在 SR 网络中验证和隔离故障的能力。IETF RFC 8287 定义了一组扩展，用于为基于 MPLS 数据平面的 SR IGP Prefix SID 和 IGP Adjacency SID 执行 LSP Ping 和 Traceroute 操作。

该测试例也是 EANTC 第一次进行 SR OAM 互操作测试。思科参与此测试例的亮点包括：

- 思科成功验证了发起 **SR OAM ping/traceroute** 操作——使用 **MPLS** 回显请求数据包，其中携带了包含新定义的 **IPv4 IGP Prefix SID sub-TLV** 的目标转发等同类栈 **TLV(FEC Stack TLV)**
- 思科成功验证了作为目标，响应 **SR OAM ping/traceroute** 操作

这次测试中，由于对 RFC8287 中 IPv4 IGP Prefix SID sub-TLV 长度存在着不同理解，部分厂商之间出现互操作问题。一家参测厂商提出了技术勘误。

EVPN

从一开始，思科就一直在 IETF 主导着 EVPN，同时在运营商和数据中心产品线的技术实现上予以落实，目前该技术已在全球运营商广泛部署。

一些人可能注意到近几年 Cisco IOS XR 产品并没有参与 EVPN 互操作测试，但今年我们全面回归，展示了 ASR 9000 和 NCS5500 产品家族先进的 EVPN 功能集。

思科参与此测试领域亮点包括：

- 思科是参与 **EVPN 测试领域的 8 个厂商之一**（不包括流量模拟器厂商）。这是有史以来参测厂商最多的一次，而且主要网络厂商均有参与
- 思科作为 **EVPN 的主 BGP 路由反射器**，所有连接到 **SR MPLS 核心网络**的参测厂商均作为其客户端
- EANTC 测试首次在核心网络中将 **SR MPLS 网络**作为 **EVPN 业务**的主要底层传送技术
- **基于 SR MPLS 的 EVPN 多归属全活测试例**
 - 多归属全活功能是 EVPN 相比传统技术（例如 VPLS）的主要优势
 - 思科成功验证了在多厂商、多归属以太网段(Ethernet Segment)环境中充当 PE
- **基于 SR MPLS 的 EVPN VPWS 测试例**

- 思科成功验证了在单归属配置中充当 PE。参测厂商同意进行多归属测试，但是时间不够了。测试期间有很多天我们是靠着超浓咖啡不眠不休地工作!
- **EVPN-VXLAN 和 IP-VPN 互通测试例**
 - 思科成功验证了通过三层 DCI，将跨 IP-VPN 广域网的多个 EVPN 数据中心互联起来

下一步?

请在 [Twitter](#) 和 [Linkedin](#) 上关注我们的最新公告。

请访问我们的[外部 SR 网站](#)以获取最新的演讲、教程和演示等等。

关于互操作测试，我期待 2019 年活动能再次取得成功。尤其是，思科对今年活动已经提出了很多重要的、基于标准的解决方案，我期待其它厂商会感兴趣，并为互操作测试做好准备，包括：

- **IGP Segment Routing 灵活算法**——Flex Algo，它是 SRTE 工具集的最新成员。Flex Algo 定义了 IGP 扩展以允许运营商自定义流量工程优化方式，并将其分配给 IGP 前缀算法。IETF 定义了 IS-IS 和 OSPF 的 Flex Algo。思科在 2018 年巴塞罗那思科用户大会(Cisco Live)上发布了这一最新解决方案，并在 [Clarence 的博客](#)和此[演示](#)中作了进一步描述
- **以 BGP 作为信令的 Segment Routing 策略**——BGP 用作对 SRTE 头端发起 SR 策略候选路径的信令。IETF 正在标准化新的 BGP SAFI 和 NLRI
- **基于 PCEP 的不相交 LSP 的信令和计算**——不相交路径是在当前 IP/MPLS 网络尤其是在二层 MPLS 传送中很常见的应用案例。其中一个 IETF 草案提出了 PCEP 扩展以发送不相交 LSP 限制条件。PCE 可以利用所期望的不相交类型，计算出满足路径计算请求的 LSP

最为重要的是，我非常期待再次在柏林度过一个愉快的冬季!