



专栏：云网融合

云网融合 PCEP 应用及端到端保障方案

王越, 王爱俊, 徐洪磊

(中国电信股份有限公司研究院, 北京 102209)

摘要: 随着云计算、大数据等新技术的蓬勃发展以及通信基础设施的不断升级, 国内三大运营商均提出以智能牵引网络转型, 构建集中控制, 灵活编排的云网基础架构的发展战略。PCEP 作为一种广泛应用于 SDN 架构的路径计算协议, 往往应用于解决复杂网络环境跨层、跨域约束路径的计算。从探讨 PCEP 在云网融合场景应用的角度出发, 提出了一种基于 PCEP 的流量端到端保障方案, 可实现在本地 IP 环境下, 面向连接的网络通信及端到端业务保障, 有助于提升运营商对网络的实时感知、敏捷响应、智能分析等能力, 满足通信资源可全局调度、能力可全面开放、架构可灵活调整的运维管理需求。

关键词: PCEP; VLAN 流量转发; 路由前缀; 业务保障

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2021203

PCEP application in cloud-network convergence and end to end traffic assurance scheme

WANG Yue, WANG Aijun, XU Honglei

Research Institute of China Telecom Co., Ltd., Beijing 102209, China

Abstract: With the booming development of new technologies such as cloud computing, big data and the continuous upgrading of communication infrastructure, the major domestic ISPs have proposed the development strategy of building a centralized control and flexible cloud network infrastructure under the trend of intelligent-traction network transformation. As a path calculation protocol widely used in the existing SDN architecture, PCEP protocol was often used to solve the problem of cross-layer and cross-domain path computing under complex network environment. The application of PCEP protocol in cloud-network convergence scenario was discussed and an end-to-end traffic guarantee scheme was introduced based on PCEP protocol which could ensure the connection-oriented network communication and make end-to-end service guarantee in native IP environment. It helps to improve the operators' ability of real-time perception, agile response and intelligent analysis of the network and meets the operation and maintenance management requirements such as global scheduling of communication resources, comprehensive opening of capabilities and flexible architecture.

Key words: PCEP, VLAN traffic forwarding, routing prefix, service guarantee

收稿日期: 2021-05-01; 修回日期: 2021-08-19

基金项目: 国资委重点研发计划项目 (No.21HQBYF0019-001); 中国电信研究院专业能力级项目 (No.T-2021-10)

Foundation Item: Key R&D Program of State Assets Administration Committee of the State Council of China (No.21HQBYF0019-001), Professional Ability Level Program of China Telecom Research Institute (No.T-2021-10)

1 引言

随着 5G、IPv6、物联网、边缘计算、云计算等新型网络和计算技术的引入,通信网络正逐渐向泛在共生、虚实共管、云网融合的新型智慧化网络发展。网络技术的快速发展,使网络呈现多业务承载、控制集中、分层开放等特性,其架构由单一业务垂直化逐渐演变为横向打通的、向业务开放的、云网融合的网络模态。其泛在化、智慧化、情境化的新型应用形态与模式改变了现有网络的部署、操作和管理方法,对网络资源的管理、融合、效能、安全等支撑能力提出挑战。其中,软件定义网络 (software defined network, SDN)、网络功能虚拟化 (network functions virtualization, NFV) 作为网络技术转型的新方向,倡导业务、控制与转发三层分离,对于网络抽象、实现网络智能管理、业务灵活调度,加速网络能力开放,增强网络集中控制与自动化管控能力起着至关重要的作用,已成为业界发展趋势。其作为运营商向“互联网+”时代转型的重要支持技术,能够使网络更加开放、通用,保证业务快速迭代和不断创新,从而推动运营商围绕云网基础资源,从云内、云间和入云到多云协同和云网边端协同,不断推进和深化。其中,IP 网络控制器作为智能网络的核心组件,能够通过实时多维拓扑感知、多约束因子路径计算、实时路径调度与重优化,实现网络闭环自动化控制,其可靠性、稳定性、安全性对网络安全运行十分重要^[1]。

2 PCEP 在云网融合场景的应用

作为当前云计算市场竞争的核心,云网融合已经成为全球领先运营商正积极推进的发展战略。云网融合不仅需要利用虚拟化技术实现传统网元设备算力、网络和存储等基础设施的集中解耦、融合和重构,保证资源的一体化供给,更要求承载网络能够根据各类云服务需求按需开放网

络能力,实现网络与云的敏捷打通、按需互联,并体现出智能化、自服务、高速等特性,实现一体化智慧运营和服务保障。从网络架构的角度出发,以 DC (data center) 为中心构建网络,重点解决云资源池和互联网数据中心的链路扁平直达问题,做到东西流量和南北流量并重。在目前的数据中心中,云资源和网络资源大多基于 OpenStack 的 Neutron 组件实现对接,功能有限,无法实现在广域网进行大量的网络配置和调整,尤其在面对混合云、跨云互联以及多点入云等场景,稍显力不从心。此外,随着网络规模的增长,对网络进行精细化管控问题也变得尤为突出。

在云网融合大趋势推进运营商向 IP 网络智能化逐步演进的背景下,控制器作为 SDN 架构的核心组件其重要性日益凸显,且从以对内应用为主,逐步转向以对外提供服务为主。作为整个网络体系的“大脑”,控制器利用北向接口使业务具备能够便利调用底层网络资源的能力,实现资源的统一调度和管理。南向接口则负责与底层设备通信,通过业务功能抽象屏蔽了底层物理转发设备的差异,实现了资源的虚拟化,完成了对底层设备的集中统一控制^[2]。

当前,在运营商的 IP 网络中,二层数据报文中只有源和目的地址的字段,是面向无连接状态的逐跳式服务,缺少针对数据流量路径的控制消息。为更好地实现流量感知和路径优化,提供应用驱动的网络服务,保障关键业务质量,均衡流量分布,运营商通常会采用多协议标签交换 (multi-protocol label switching, MPLS) 网络或者 IPv6 网络的方案。随着业务量的增加,网络规模不断扩大,这样的做法将过多的复杂协议引入封闭的网络设备,网络设备不仅需要承担数据转发、路由管理、路由协议解析等任务,还要消耗大量计算资源进行复杂约束条件的路由运算^[3],给网络基础架构带来额外负担的同时,增加了管理难度和运维工作量。



PCEP (path computation element communication protocol, 路径计算单元通信协议) 作为基于 TCP 的路由集中控制方案, 是由 IETF 的 PCE (path computation element, 路径计算单元) 工作组于 2006 年为 MPLS 网络域间流量工程等应用提出的通信协议^[4]。其最初应用于光网络, 随着南向接口技术在 SDN 体系发展的日趋成熟, PCEP 也不断扩充和完善, 逐步实现了面向分段路由的以及 LSP 保护路径的扩展等能力^[5]。作为众多南向接口协议 (如 BGP LS、OpenFlow、Netconf) 中较为成熟的 PCEP, 其采用了集中式的路径计算模型, 该协议将路由器的 CSPF (constrained shortest path first) 功能抽离, 通过集中部署控制器或者 PCE, 根据全网的带宽、代价、标签等全局资源视图进行约束路径计算, 以集中算路的方式, 为各类应用计算最优路径, 下发给网络设备, 控制报文转发, 实现全局统一调度, 完成路径计算和路径建立转发功能的分离^[6], 保障流量端到端全局最优, 是一种应用较为广泛的网络路径计算协议。相比于分布式路径计算, 通过 PCE 实现集中算路的方式在解决复杂网络环境跨层、跨域的端到端约束路由计算方面更具优势, 同时能够有效降低运维复杂度与难度, 为简化网络运维, 提升网络质量提供了可能。

3 基于 PCEP 协议的流量端到端保障方案

3.1 端到端保障方案需求

在路由管控方面, 传统的 MPLS-VPN 在原有 IGP (interior gateway protocol) 基础上增加 LDP (label distribution protocol) 实现标签的转发过程, 由于 LDP 不具备流量工程, 而通过 RSVP-TE (resource reservation protocol-traffic engineering) 实现显示路径计算信令较为复杂, 同时庞大的 TE 链路信息扩展和维护困难, 设备协议开销以及运维复杂度较高, 无法实时采集监控网络质量情况并完成最优路径的自动优化, 因此信息交互效率

低。作为新兴技术的 SRv6, 虽充分利用 IPv6 扩展头的机制, 通过 SRH (segment routing header) 中的 IPv6 地址标识 segment 实现流量的引导转发, 但因 SRH 信息的引入, 报文开销较大、网络链路带宽利用率低。此外 SRv6 报文处理对芯片要求较高。现网设备芯片难以支持对 128 bit SID (segment ID) 的扩展头 SRH 复制和操作, 无法实现向 SRv6 的平滑升级演进, 给运营商部署 SRv6 带来较大成本压力。

考虑到 MPLS 以及 SRv6 类型方案要么利用 L3 VPN 技术, 要么利用 SRH 扩展 IPv6 包头, 对于 IPv4 网络, 缺乏良好的解决方案。针对 IP 网络中无连接的状态, 本文利用 PCEP 转控分离的机制, 提出了一种基于 PCEP 的流量端到端保障方案。该方案有效利用了二层以太帧结构中的 VLAN 信息, 通过 PCE 实时收集全网资源信息, 响应转发节点请求, 完成智能路径计算和头端计算路径自动下发, 从而保证在本地 IP 环境实现面向连接的网络通信及端到端业务保障^[7]。相较于 MPLS、SRv6 等方案, 本方案利用了全新的 VLAN 地址空间, 避免了与其他已有协议的冲突, 能够同时适用于 IPv4 和 IPv6 网络。

通过 PCE 的全域资源感知, 能够获得云网资源的一致质量保证, 确保业务的一体化规划和运维管理, 实现云业务和网络业务的深度融合。结合 Telemetry 和 Netflow 等技术亦可实现业务流量可视化和按需调度, 保证网络资源和云资源的一体化弹性供给和敏捷服务。

3.2 端到端保障方案模型架构

基于 PCEP 的流量下发保障方案, 主要思路是控制器根据网络约束参数完成关键数据业务端到端的初始化路径计算, 并通过和设备建立 PCEP 将结果以三元组和交叉表方式通过南向接口下发至设备。设备收到信息, 在本地形成路由映射表和路由交叉表, 根据表项内容, 将数据流量封装对应的 VLAN_ID 进行转发, 实现数据流量的二层封装

和面向连接的快速转发,保障业务端到端质量。

本方案系统主要包括 SDN 控制器、边缘路由设备(入口路由器和出口路由器)、转发路由设备。

其中,SDN 控制器通过 PCEP 分别与系统边缘设备、转发路由设备建立连接,根据实际业务需求,完成业务端到端保障路径的计算,形成三元组信息和网络交叉信息分别下发至边缘和转发设备;边缘入方向设备(入口路由器)负责向控制器发送路径请求初始化确认信息,同时根据控制器下发的路径计算结果在本地形成路由映射表,设备依据此表对接收的数据报文进行匹配,对匹配的报文封装对应的 VLAN-ID 进行转发;转发路由设备根据控制器下发的路由交叉表,完成从入方向子接口标识到出方向子接口 VLAN-ID 标识的交叉转发和数据再封装。而边缘出方向设备(出口路由器)则将 VLAN-ID 解封套,并基于数据目的地址进行转发。

在实际业务场景中,可以在网络拓扑中边缘入方向设备和出方向设备分别部署多个 BGP session,不同的 BGP session 分发不同的前缀,并具有不同的 BGP 下一跳。在控制层,入口路由设备通过 BGP 可以学习网络拓扑中由控制器约束的基于同一 BGP session 的源/目的 peer 的不同路径,即源和目的 BGP 前缀。控制器作为 PCE 则根据网络拓扑、流量工程策略(TE policy)等参数信息计算满足约束条件请求的路径,通过 PCEP 的 PC 初始信息将路径通过三元组信息——源 BGP peer、目的 BGP peer、VLAN-ID 下发至作为路径计算客户端(path computation client, PCC)的入口路由设备,完成 VLAN 路径的初始化。入口路由设备收到 PCEP 下发的三元组信息后,向控制器发送 PCRpt 消息完成确认,同时形成基于源/目的 BGP 前缀的映射路由表,最终转化为实际的隧道和隧道路径的配置。

在转发层,当入口路由设备接收到数据报文时,会将数据报文的源/目的地址和映射路由表中

源 BGP 前缀、目的 BGP 前缀进行匹配,如果一致,则将数据包打上对应的 VLAN-ID 标签,并创建对应的 VLAN 子接口进行转发。

转发路由设备收到封装对应 VLAN-ID 的数据报文后,通过匹配控制器下发的 VLAN 交叉表,将数据报文原有 VLAN 信息解封套,打上新的 VLAN-ID 标签,同时创建对应的 VLAN 子接口进行转发。

出口路由设备则根据映射信息,将封装 VLAN-ID 的二层包头去掉,对报文进行三层转发。实现了在本地 IP 环境下,面向连接的网络通信及端到端业务保障。解决了传统 IP 报文无连接无状态的痛点。

3.3 基于 PCEP 架构的 VLAN 流量转发流程

对于现网中较为复杂的网络拓扑,网络设备之间往往会存在多个 BGP session,不同的 session 用于承载不同的数据业务流,用于实现流量的端到端隔离和不同级别的 SLA(service level agreement)等级保障。基于前述端到端保障方案模型架构设计,PCEP 架构的 VLAN 流量转发方案的实施流程如图 1 所示。

在 R1 和 R8 之间通过不同的子接口创建 3 个 BGP 实例,分别对应 3 个 BGP session,不同的 BGP session 分发不同的前缀,并具有不同的 BGP 下一跳。

若要对 R1 和 R8 之间的某些关键业务进行流量端到端保障。在 BGP session 1 中业务流量的源 IP 前缀(source IP-prefix)和目的 IP 前缀(destination IP-prefix)分别为 P1 和 P8,入口设备通过 BGP 可以学习到网络拓扑中基于 BGP session 1 的源/目的 peer 的不同路径前缀 P1、P8。

控制器和入口设备 R1 通过 PCEP 建立连接。控制器根据全局网络拓扑、TE policy 等参数信息计算全局优化路径,形成基于 BGP session 1 的三元组信息(BS1_R1、BS1_R8、VLAN12),其中,BS1_R1 和 BS1_R8 分别表示基于 BGP session 1 的 R1 和 R8 的逻辑子接口地址,VLAN12 表示

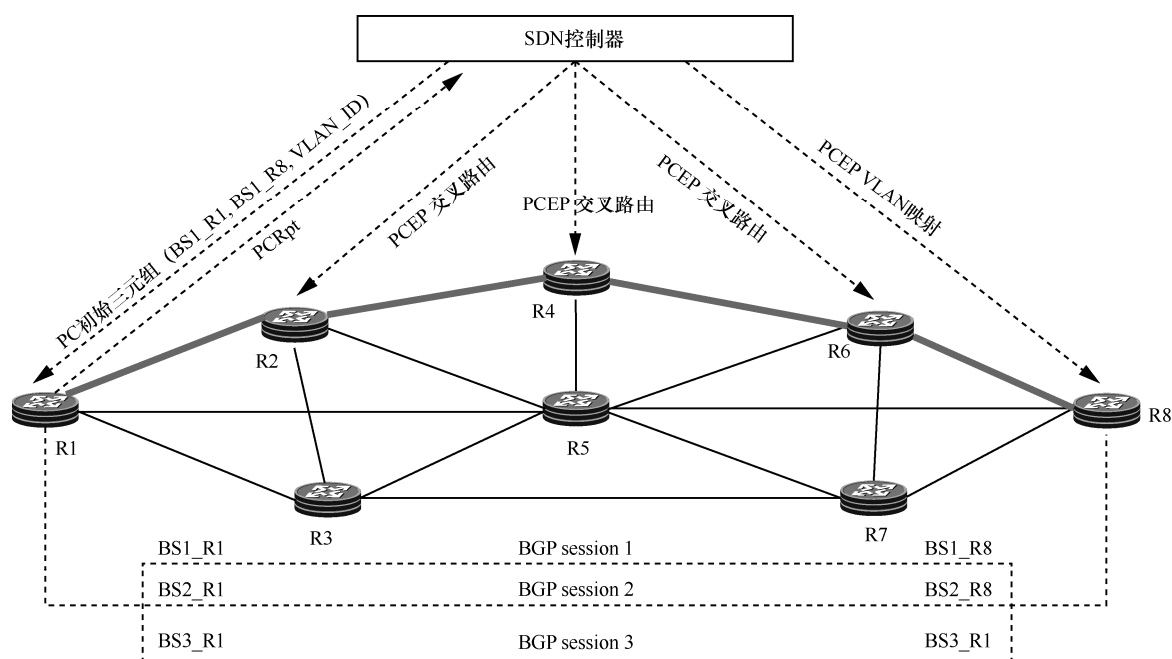


图1 基于 PCEP 协议的流量端到端保障方案

R1 到 R2 路径所对应的 VLAN。控制器将该三元组信息通过 PCEP 下发至 R1, R1 收到信息后返回 PCRpt 确认信息。

入口设备 R1 根据收到的三元组信息, 结合之前学到的路径信息形成基于源/目的 BGP 前缀的映射路由表, 示例见表 1。

表 1 映射路径表

P1 前缀	P8 前缀	VLAN
$\begin{pmatrix} \text{Src-prefix1} \\ \dots \\ \text{Src-prefixN} \end{pmatrix}$	$\begin{pmatrix} \text{Dst-prefix1} \\ \dots \\ \text{Dst-prefixM} \end{pmatrix}$	VLAN12

其中, P1 前缀和 P8 前缀分别对应 session i 中业务流量的 source IP-prefix 和 destination IP-prefix, P1 前缀匹配 N 条源 IP, P8 前缀匹配 M 条目的 IP, 因此该路由表针对 session 1 共有 $N \times M$ 条路径组合。

当入口设备接收到数据报文后, 会查找映射路由表, 需要保障的报文通过映射路由表的匹配被打上 VLAN12 标签, 送至对应 BGP session 1 的

逻辑子接口进行转发。

对于中间设备 R2、R4、R6, 控制器会下发交叉路由表。交叉路由表中的条目为 VLAN 键值对, 对于 R2, 条目是 (VLAN12、VLAN24), 当收到封装对应 VLAN-ID 的数据报文后, 中间设备 R2 会将报文解封装, 去掉由 R1 打上的 VLAN12 标签, 打上 VLAN24 标签送至 R4, 同时 R4 创建对应的 VLAN 子接口进行转发, R4 和 R6 的操作同理。

出口设备 R8 收到 SDN 控制器下发的交叉路由表, 当 R8 收到 R6 送来的报文, 通过匹配路由表得知自己是最后一跳目的地址, 则不进行任何封装, 直接将 VLAN-ID 的二层包头去掉, 对报文进行三层转发。

至此, 对于从 R1 到 R8 的数据报文, 只有匹配到 P1 和 P8 前缀的数据业务, 才能使用端到端的逻辑通道, 并且该逻辑通道的路径是可由 SDN 控制器通过 PCEP 下发特定路由信息进行规划的, 对于其他接口进入的数据, 则依据传统的路由表进行转发, 实现了在本地 IP 环境下, 面向连接的网络通信及端到端业务保障的诉求。

以太接口的大规模部署,使得利用二层帧结构中包含的信息简化端到端数据转发成为可能。PCEP 架构基于 VLAN 的流量转发,充分利用了二层 VLAN 信息,以 PCEP 为基础实现了流量的全场景接入,在尽可能保留 PCEP 结构的同时简化了报文的端到端路径计算和转发过程,能够满足多业务流量不同路径转发需求,确保关键业务的优先级和服务质量,从而实现灵活组网以及多维度 SLA 路径规划。

4 结束语

本文提出的基于 PCEP 的流量端到端保障方案,可针对特定客户、特定应用进行端到端业务保障,实现 IP 场景面向关键业务的确定性传输,相较于当前主流的 MPLS 和 SRv6 的解决方案,可以减少标签资源的占用,减轻标签转发表的维护工作量,提高转发效率,提高整个网络传输效率。借助网络云虚拟化、白盒、云原生等核心能力落地,可实现业务的快速开通和确定性转发。本方案利用顶层控制器完成端到端的流量和业务管理,能够保证全网资源的优化调度和协同编排,有助于构建云网资源一体化管控的云网操作系统,完成云网资源的统一抽象、统一管理、统一编排、统一优化,实现云网融合大环境的网络智慧化运营,助力构建简洁、敏捷、开放、融合、安全、智能的云网生态系统。

参考文献:

- [1] 卢泉,梁筱斌,尹远阳. IP 网络控制器架构与关键技术[J]. 移动通信, 2020(9): 62-67.
LU Q, LIANG X B, YI Y Y. Architecture and key technologies of IP network controller[J]. Mobile Communications, 2020(9): 62-67.
- [2] 徐洪磊,孙琼. 基于 SDN 的网络调优及控制器北向接口探讨[J]. 电信技术, 2016(6): 49-52.
XU H L, SUN Q. Discussion on network optimization and controller north interface based on SDN [J]. Telecommunications Technology, 2016(6): 49-52.
- [3] 唐宏,刘汉江,陈前锋. OpenDaylight 应用指南[M]. 北京: 人民邮电出版社, 2016.
TAGN H, LIU H J, CHEN Q F. OpenDaylight Application Guide[M]. Beijing: Posts and Telecom Press, 2016.
- [4] 鞠卫国,王跃庆,林文祥,等. SDN 南向接口协议探究[J]. 电信快报, 2018(1): 8-12.
JU W G, WANG Y Q, LIN W X, et al. Research on SDN southbound interface protocol [J]. Telecommunications Information, 2018(1): 8-12.
- [5] 周福博. 软件定义光网络南向接口 PCEP 的设计与实现[D]. 北京: 北京邮电大学, 2020.
ZHOU F B. Design and implementation of software defined optical network south interface PCEP[D]. Beijing: Beijing University of Posts and Telecommunications, 2020.
- [6] IETF. Path computation element (PCE) communication protocol (PCEP): RFC5440[S]. 2009.
- [7] IETF. PCEP Extension for Native IP Network: draft-ietf-pce-pcep-extension-native-ip-16[S]. 2021.

[作者简介]



王越(1986—),男,中国电信股份有限公司研究院工程师,主要研究方向为 IP 网络相关技术、SDN 技术及 IETF 相关标准(BGP/PCE 等)。



王爱俊(1973—),男,中国电信股份有限公司研究院高级工程师,主要研究方向为 IP 网络技术及 IETF 相关标准(BGP/IGP/PCE 等)。



徐洪磊(1982—),男,中国电信股份有限公司研究院高级工程师、SDN 解决方案研究室主任,主要研究方向为基于 SDN 的 TE 优化技术、SD-WAN 等。