



SDN 在 IP 广域网的典型应用与实现

黄勇军, 陆小铭, 曹维华

(中国电信股份有限公司广州研究院, 广东 广州 510630)

摘要: SDN 最重要的 3 个特征是转发控制分离、集中控制和开放接口, 现阶段 SDN 的产品和系统尚未成熟, 运营商通过 RR+、PCE 等技术逐步实现 IP 网络的集中控制功能, 是传统 IP 网络向 SDN 演进的有效途径。首先分析了 SDN 在广域网的应用系统的架构和各层的主要功能, 并对其典型应用, 如路由优化调度、流量均衡调度及智能化路径管理等进行了说明。

关键词: SDN; RR+; BGP; PCEP

中图分类号: TN915

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016111

Typical applications and realization of SDN in IP WAN

HUANG Yongjun, LU Xiaoming, CAO Weihua

Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China

Abstract: Three key points of SDN are decoupling of data plane and control plane, centralized control and open interfaces. Some carriers develop a SDN application system based on RR+ or PCE technology in order to achieve the centralized control function on traditional IP WAN. The architecture and development suggestions of the SDN application system were presented, and some typical functions, such as routing optimizing, traffic balanced scheduling, paths intelligent arrangement and so on, were described.

Key words: SDN, RR+, BGP, PCEP

1 引言

当前, SDN 正逐步实现从理论到标准、从技术到产品、从部署到服务的过程, 业界从数据中心、骨干网、接入网和移动核心网多个应用场景研究 SDN 的应用和实践^[1]。从 IP 广域网的角度看, 传统网络在流量调控、路由优化、业务差异化调度等方面面临挑战, 传统分布式路由计算的网络架构难以满足日益增长的网络业务和质量要求。SDN 技术提出了集中的网络管控模式, 将传统网络架构分拆为应用、控制、转发分离的架构, 从而增强对网络业务的调度管控能力。作为全新的网络架构技术, SDN 的主要特征^[2]包括以

下几个方面。

(1) 控制平面和数据平面分离

控制平面独立, 采用通用硬件处理实现; 控制平面和数据平面采用标准接口(如 OpenFlow), 通用硬件只完成必要的转发功能。

(2) 网络智能和状态逻辑上集中化

业务逻辑、路由计算、转发表项等实现集中管理和调度。

(3) 从应用的角度看, 下层网络被抽象出来

网络具备可编程性, 可实现自动化管理; 网络更加灵活, 具备更强的扩展性。

SDN 强调通过软件的方式来定义网络的行为, 提升网



络控制调度能力,为网络优化提供了很好的思想和手段,其架构具有普遍适用性。但是,SDN在技术标准、系统和设备实现方面面临成熟度需提升的问题。SDN的应用一方面要考虑技术的可行性,即现有网络的支持度;另一方面要考虑现实的应用场景,从运营需求和用户需求出发逐步试点和推进。

2 SDN在IP广域网的系统实现

传统IP网络主要采用IGP(interior gateway protocol,内部网关协议)作为内部路由的计算,采用BGP(border gateway protocol,边界网关协议)实现网络自治系统间的路由通告,路由作为流量转发的关键依据,一直是网络管理者关注的重点。对于SDN在IP广域网上的实现,业界有专家提出了“传统网络→RR+→PCE+→SDN”的演进方法和步骤^[3],从而较好地保障了网络的原有投资,实现了网络的平滑升级。在第一阶段,SDN在IP广域网的应用系统主要是在传统网络叠加增强的路由反射(RR+)模块,围绕全网路由的分析及优化,同时结合流量管理、业务质量检测等能力,在一定程度上实现传统网络的集中路由、路径

和流量的控制;第二阶段,网络设备需要支持PCEP的功能,对流量在网络内部的路径进行精细控制,提供多种控制手段和策略,完全控制隧道的路径计算和建立^[4]。

从现网情况看,传统的IP广域网都建设了相应的网管系统,网管系统经过长时间的迭代开发,除了传统的FCAPS(故障、配置、计费、性能、安全)管理能力,还根据IP网络的特点拓展了业务管理、维护管理、综合分析等功能模块,实现了对IGP/BGP路由的全局管理、流量流向分析、VPN业务自动开通和业务管理等功能。SDN的业务编排及网络控制模块可以添加到网管系统中,充分利用已有的拓扑、路由、流量、质量测试等维护管理能力。图1是其中一种SDN在IP广域网的应用系统的系统架构。

该系统功能架构基于软件的分层思路,整个架构划分为应用层、产品管理、业务编排、网络控制、运维管理(包括业务管理和网络管理)、网络数据库等,以虚拟化与接口插件化为基础实现各部分间的融合交互和灵活调用。各层的主要功能如下。

应用层处于架构的最上层,是面向用户的核心价值所在,通过开放API,支持第三方、用户自定义、内部运维管理。

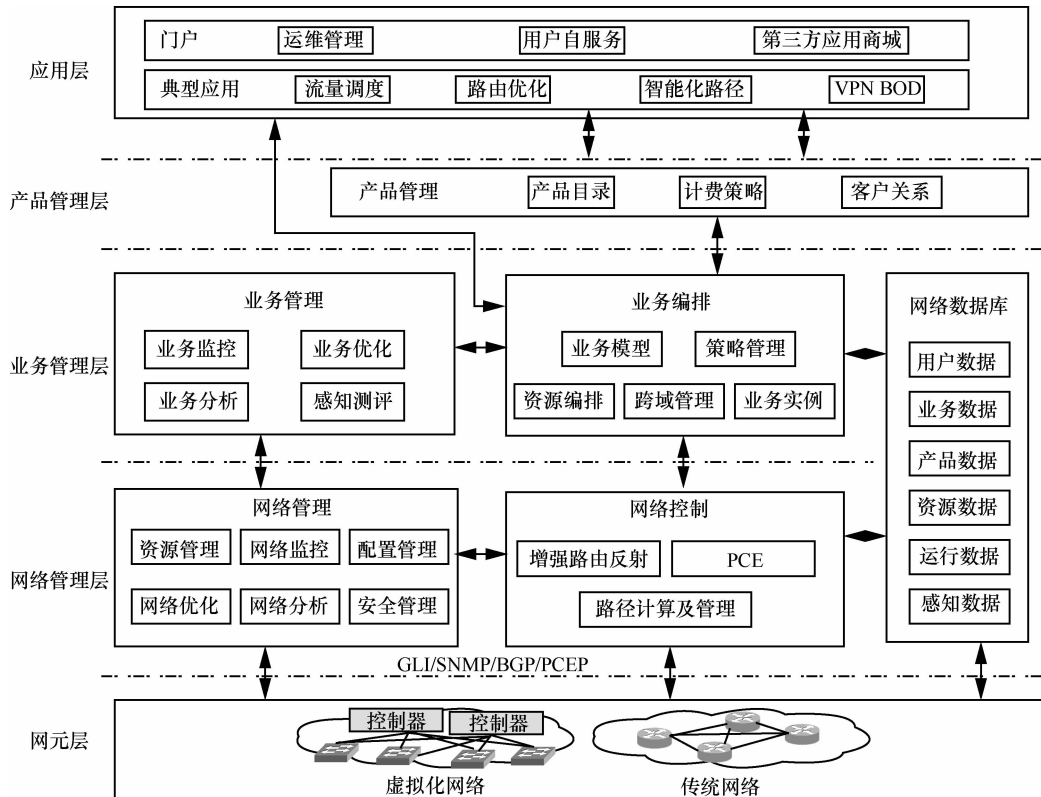


图1 SDN在IP广域网的应用系统功能架构

(1)产品管理

以产品形态满足服务需求,提供产品目录、产品规格、定价策略、产品套餐,把应用变成可销售的产品,从而加速网络业务创新从创意到上线。

(2)业务编排

基于业务约束条件,通过模型驱动和跨域资源协同将各种网络服务单元基于各种策略进行有序地安排和组织,提供差异化的网络业务创新。

(3)网络控制

将底层网络资源抽象成可操作的信息模型提供给上层,并将上层的网络需求转化为底层的网络控制指令,下发到网络设备中,从而封装并开放网络能力,屏蔽南向差异,实现网络按需控制。

(4)运维管理

聚焦业务管理和网络管理,包括监控基础设施、控制器及整体架构涉及的网元运行情况,并利用 SLA、QoE 等技术,实现业务质量与网络运行的协同感知。

(5)网络数据库

面向产品、业务和资源,采集和存储网元的运行信息、网络资源信息及业务质量信息等,支持对日志、告警、信令等非结构化数据的分布式存储和数据挖掘,实现数据的共享和分析。

SDN 在广域网的典型应用中,增强的路由管理模块是实现路由层面集中智能控制的关键,包含以下能力。

(1)IGP 路由管理

通过建立 ISIS 的“哑邻居”的方式,获取全网的 IGP 路由,能描绘 IGP 路由拓扑,对 IGP 路由波动进行监控,对存在多条链路(IGP 连接)的端口 IP 地址进行分析等。

(2)BGP 路由管理

通过建立“哑”peer 方式,实时获取 BGP 路由信息,实现对 BGP 路由事件的实时监控,实时路由表及路由变化的查询、路由告警功能等。

(3)RR+功能

RR+能力是基于传统 RR 能力的扩展功能,可根据业务编排模块的要求向指定路由器反射指定的 BGP 路由信息,如改写下一跳(next-hop)、调高路由 local preference 值、添加 community 标记等操作,从而引导流量按照网络管理者的要求进行调度。

3 SDN 在 IP 广域网的典型应用

3.1 路由优化调度

路由优化调度的一个重点应用是解决 IP 广域网中传统 RR 在路由反射中存在的缺陷。传统 RR 是从邻居节点收到路由后,按照 BGP 选路规则优选其一,然后把优选的路由发送给其他所有邻居节点。优选规则中的 IGP 开销值是 RR 本身到 BGP 下一跳(通常是路由发布者)的值,这个规则在实际网络中会带来路由不优的问题,如图 2 所示。

具体分析如下。

- R1 和 carrier A 相连,R6 和 carrier B 相连,刚好 carrier A 和 carrier B 都通告 prefix 1 的 BGP 路由。
- 传统 RR 收到 R1 和 R6 通告的 prefix 1 的路由,根据选路规则,从 IGP 开销的角度看,RR 到 R1 的开销要比 RR 到 R6 的小,RR 选择到 R1 为最优路由向所有 iBGP 邻居进行反射。
- AS 3 去往 prefix 1 的流量路径是 R3→R2→R1→(prefix 1),为最优路径。

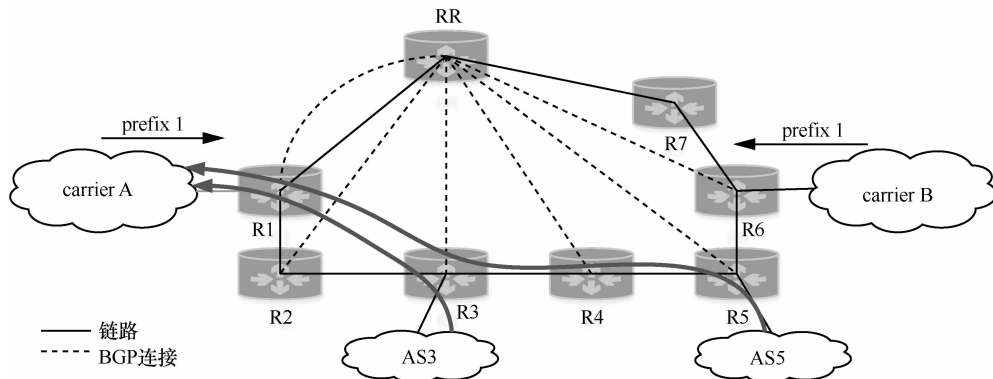


图2 传统RR选路存在问题的示意图



- AS5 去往 prefix 1 的流量路径:R5→R4→R3→R2→R1→(prefix 1),不是最优路径。

由于现实网络中常常会存在相同路由多出口的情况,传统 RR 的选路会导致次优路由产生,流量会走“长路径”,从而影响业务质量。因此需要实现基于“源目的”的路由反射增强机制,其中基于源目的 IGP 最短路径的 BGP 路由反射是较为有效的方法,其实现原理如图 3 所示。

主要实现原理如下所述。

(1)RR+部署

在网络中叠加 RR+路由管理模块,RR+从传统 RR 获取全网 BGP 路由,同时和 R1~R6 等设备建立 BGP 邻居关系,R1~R6 向 RR+发送 best-external 路由。同时,RR+可以向 R3、R5 反射指定的路由。

(2)多出口路由信息计算

RR+对接收到的全部路由进行归并计算,找出有多个出口(next-hop)且 BGP 选路属性(local 值、AS path 长度、MED)相同的 prefix 及其路由集合(这些路由需要通过 IGP metric 进行路由优选)。

(3)源和出口间的 IGP metric 计算

IGP 路由分析模块在采集全网 IGP 路由的基础上,计算源(R3、R5)与各出口(R1、R6)之间的 IGP metric。

(4)RR+优选路由反射

结合源与各出口之间 IGP metric,选择路径最短的一个出口作为主选路径,通过管理员确认或者自动方式,向 R3、R5 分别反射最优路由,该最优路径的 local preference 值比传统 RR 反射的路由要高。

(5)R3、R5 接收到 RR+反射的优化后的路由,自动适配为最优的路由路径,此时,AS 3 去往 prefix 1 的流量路

径为 R3→R2→R1→(prefix 1),AS5 去往 prefix 1 的流量路径:R5→R6→(prefix 1)。

因此,系统通过 RR+及 IGP 最短路径相结合的优化算法,可以有效规避传统 RR 选路的缺陷,提升网络的转发效率和用户体验。此外,还需要实现 RR+模块的路由撤销和不断检查机制,如系统应能定时分析现存的多路径路由,如果发现增加,向源路由器反射路由;如果发现减少,可撤销所反射的路由。

3.2 流量均衡调度

在 IP 广域网上,由于用户的分布不均、业务特发性等原因经常存在局部流量不均衡的现象。如图 4 所示,网络片区 B 的流量流入网络片区 A,流量分担在链路 1~3 上。流量的目的端可能是各个城域网或 IDC,由于其用户规模或业务存在着巨大差异,这几条电路的流量容易出现不均的情况,情况严重时部分电路已经发生拥塞而部分电路还处于轻载。这类现象容易出现在国际互联电路、国内运营商间的互联互通电路以及国内骨干网的省际电路等场景。

网络局部流量不均的情况涉及路由、流量和业务层面等诸多复杂的因素,同时流量动态变化快,要进行人工干预基本不可行,因此一直是困扰运营商的难题。

RR+的出现使路由层面的集中管理变成可能,网络管理者可以在流量可视化的基础上,有针对性地改变指向片区 B 的部分路由,如修改 iBGP 路由的 next-hop 参数等,从而引导部分流量,实现这几条链路上的流量均衡。

基本流量调度功能步骤如下所述。

(1)流量流向可视化分析

通过 NetFlow 分析各条分担电路的流量流向情况,展

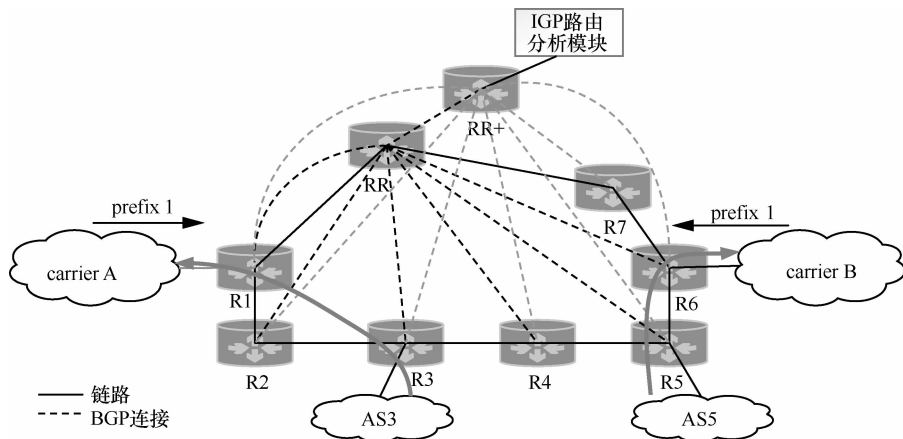


图 3 IP 广域网流量不均分析示意

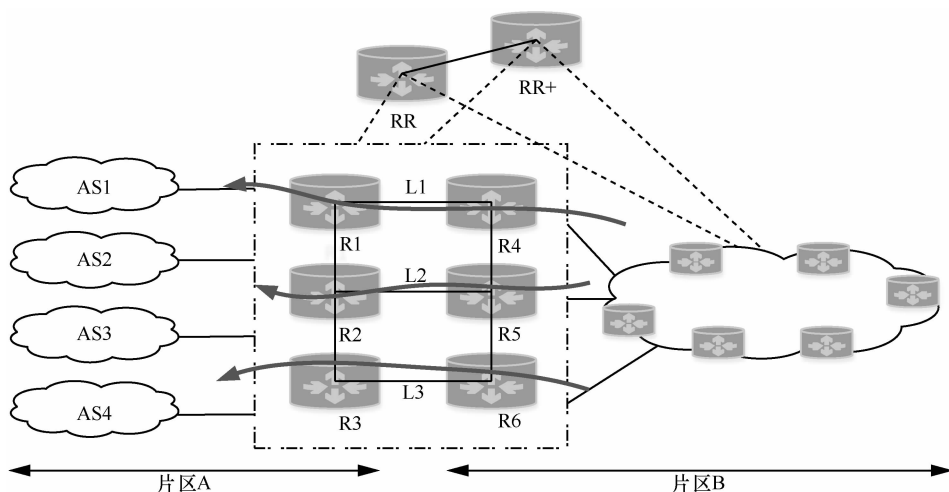


图4 基于源目的 IGP 最短路径的 BGP 路由优化示意

现以城域网/IDC 为单位的流量流向 TopN 分析数据,包括链路组、详细链路的日/月流量波动图以及流入各个城域网/IDC 的流速以及占比,为流量调度提供基础数据和调度依据。

(2) 预设流量调度策略

网络管理者将流量调度的要求(如调度的流量流向、调度方式、调度时间等)转化为路由策略,即指定目标地址的 iBGP 路由将发送到网络 B 片区的路由器上,这些路由的 next-hop 参数能够指定,同时 local preference 值比传统 RR 反射的路由大。

(3) 网络仿真

在定制的流量调度策略实施前,可先进行网络仿真,展现中继链路利用率发生变化的 TopN 链路,若链路利用率超过预设的门限将产生告警,用户可选放弃策略的调度。若中继链路利用率不越限,可通过用户确认,开始实施流量调度。

(4) 流量调度策略执行

通过 RR+反射定制的 iBGP 路由(根据 AS、community 属性匹配目标网络路由)到网络 B 部分的路由器,通过改变路由引导流量走向。

(5) 分析功能

- 系统能输出策略应用分析报告,内容包括如下方面:
- 策略名称、策略状态(启用、停用、停用原因等)、创建时间、启用时间、停用时间、策略执行消耗时间等;
 - AS 号、流入城域网或 IDC、设定的出口、规划的出口;
 - 调度对象总流速、经设定出口流速;
 - 提供链接查看规划出口以及设定出口链路组、详细

链路调度前后流速对比、时延、分组丢失率等网络质量对比。

以上是流量均衡调度的基本步骤,在系统实现中还希望能做到自动化的均衡能力,要解决的几个关键问题如下所述。

(1) 划分可调度区域

需要在系统中指定哪些目标地址段的流量可以被调度,可以根据 AS 号或 community 进行划分。

(2) 建立流量调度数据库

流量调度数据库用于维护被调整的 AS 列表及相关必要信息,作为流量仿真及流量调度的依据。该数据库用以指导 RR+如何进行路由的集中控制。

(3) 自动化调度的触发和优化

网络管理者可以根据需要设定自动化调度的触发条件,一种较为方便的方法是通过计算标准差的方式,如果利用率的离散度比较大,可以实施调度,离散度变小,意味着调度达到了成效。

(4) 调度路由的新增和撤销

新增时,RR+根据 AS、community 属性匹配目标网络路由,修改路由的下一跳为新出口,向源路由器反射路由。需要撤销路由时,RR+根据 AS、community 属性查找已反射的目标网络路由,向源路由器发送该路由的 withdraw 消息。

3.3 智能化路径管理

运营商的 IP 广域网与其他多个运营商存在互联互通的关系,同一个用户(相同的路由)可能会通过不同的互联互通运营商通告进来,存在多条出口路径情况,如图 5 所示。

用户位于 AS1 中,要访问远端的服务器,由于 carrier A 和 carrier B 都通过 eBGP 宣告了 prefix 1 的路由,因此存

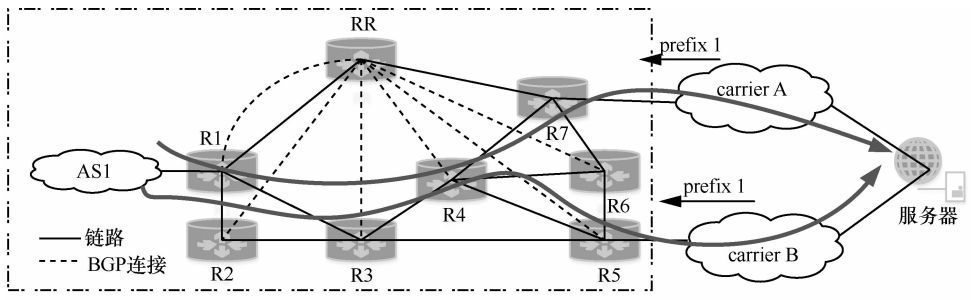


图5 用户多出口路径情况示意

在多条出口路径的情况,一条路径是 $AS1 \rightarrow \dots \rightarrow R5 \rightarrow \text{carrier B} \rightarrow \text{服务器}$;另一条路径是 $AS1 \rightarrow \dots \rightarrow R7 \rightarrow \text{carrier A} \rightarrow \text{服务器}$ 。这两条路径的服务质量会发生波动,某些时段 $R5$ 出口质量最优, $R7$ 出口较差,在某些时段则相反。传统网络中 RR 的选路相对比较固定,并不会根据业务的质量情况进行调整。

为了给客户提供长期质量保障服务,SDN 近期应用系统可以实现基于业务质量的路径管理机制,其原理如图6所示。

实现步骤如下所述。

(1)分析用户路由的多路径信息

根据用户访问目标服务器的IP地址信息,通过收集边缘业务路由器的路由,获取服务器的所有出口路由器。

(2)业务质量测试

业务质量测试系统从出口路由器($R5$ 、 $R7$)分别发起到服务器的ping测试或业务质量测试(如HTTP测试),从靠近用户的路由器($R1$)发起到出口路由器($R5$ 、 $R7$)的ping测试,同时通过位于 $AS1$ 内的探针发起到服务器的业务质量测试(用于对实施效果的评估依据)。

(3)路径质量评估

根据质量测试结果,分析不同的两条路径的质量优

劣,如把分段测试的时延相加进行比较,同时可设定一个阈值,如两条路径的时延偏差小于10%则不做处理,以避免引起用户访问路径的频繁波动。

(4)路径优化调度

当系统发现用户的当前主用出口质量劣化,同时两条路径的质量差别超过一定阈值,系统即可实施路径优化调度,通过CLI将指令发到备用边界路由器,将备用出口的路由优先级提高(提高local preference值),从而将流量引导到质量更好的备用出口。

4 结束语

本文提出的SDN在IP广域网的应用系统,充分利用了现有IP网络的资源和能力,对现网的改造小且易于实施,适应了SDN技术融合发展趋势。系统在实际应用中有效提升了网络资源利用效率和业务感知,为用户差异化服务提供了新的思路和方法。由于系统需要对网络的整体和细节有通盘的了解,因此需要与原有网管系统紧密关联,后期需要在自动化调度能力上加以完善,才能具备灵活商用的能力。未来随着网络设备逐步支持PCEP、OpenFlow、segment routing等技术和协议,逐步实现控制转发分离,网

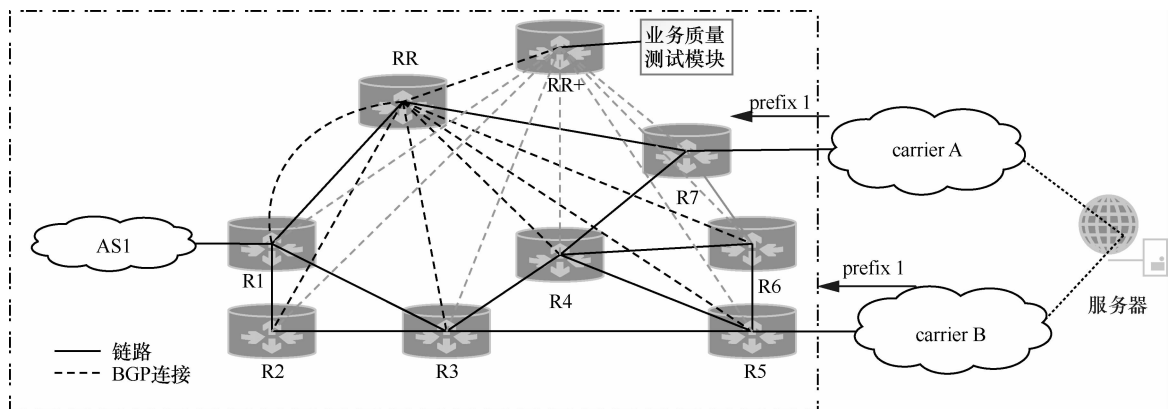


图6 基于业务质量的路径管理示意

络的集约化和智能化水平将逐步得到提升。

参考文献:

- [1] 王茜, 赵慧玲, 解云鹏. SDN 标准化和应用场景探讨[J]. 中兴通讯技术, 2013, 19(5): 2-5, 21.
WANG Q, ZHAO H L, XIE Y P. Standardization and deployment of SDN[J]. ZTE Technology Journal, 2013, 19(5): 2-5, 21.
- [2] 张朝昆, 崔勇, 唐嵩嵩, 等. 软件定义网络(SDN)研究进展[J]. 软件学报, 2015, 26(1): 62-81.
ZHANG C K, CUI Y, TANG H H, et al. State-of-the-art survey on software-defined networking (SDN)[J]. Journal of Software, 2015, 26(1): 62-81.
- [3] 王素彬, 郭亮. ChinaNet 向 SDN 演进的方法探索 [J]. 电信科学, 2013, 29(6): 151-157.
WANG S B, GUO L. Study on IP RAN technique evolution and network construction [J]. Telecommunications Science, 2013, 29(6): 151-157.
- [4] Path computation element (PCE) communication protocol (PCEP): RFC5440: 2011 [S/OL]. [2011-01-08]. RFC5440. <http://wenku.baidu.com/link?url=M7rA7byGYBnJT3GIXS7Y53I3AJCnUfKlJozdbYfjmFNTj671yXYg0RjRKi4ZcVGdKUYOTrfA8RmB-vMM745sSPy-dQjtsOh2yTfelSIRe70>.

[作者简介]



黄勇军(1970-),男,中国电信股份有限公司广州研究院副院长、高级工程师,长期从事通信网络发展规划、新技术应用研究和投资管理工作,主要研究方向为宽带网络技术、SDN 及云计算。



陆小铭(1979-),男,现就职于中国电信股份有限公司广州研究院,主要从事 IP 网络技术和 SDN 方向的技术研究工作。



曹维华(1976-),女,现就职于中国电信股份有限公司广州研究院,主要从事 IP 网络技术、支撑系统、网络分析、移动互联网、SDN 等领域的开展研发和支撑工作。