

SDN 规模部署关键问题分析

Key Issues in Large-Scale Deployment of SDN

李晨/LI Chen
程伟强/CHENG Weiqiang
王金柱/WANG Jinzhu

(中国移动通信有限公司研究院, 北京
100058)
(China Mobile Research Institute, Beijing
100058, China)

中图分类号: TN929.5 文献标志码: A 文章编号: 1009-6868 (2016) 06-0031-005

摘要: SDN 是中国移动未来网络项目 NovoNet 的核心技术之一。指出中国移动的 SDN 实践, 重点包括数据中心、软件定义的分组传送网(SPTN)、广域网、集客网、移动核心网五大应用场景。通过对 SDN 发展的关键技术问题的介绍和分析, 认为 SDN 把软件和网络结合在一起, 给网络带来灵活性和开放性, 同时也对网络固有的稳定性有所冲击。对于 SDN 引入来说, 应该全局考虑业务系统对稳定性和灵活性的需求和平衡。

关键词: SDN; NovoNet; 数据中心; SPTN

Abstract: Soft defined network (SDN) is one of the core technologies in the NovoNet project of China Mobile. The major scenarios of SDN practice of China Mobile are including data center, defined packet transport network (SPTN), wide area network (WAN), NovoVPN and mobile core network. It is considered that combing the software with network, SDN brings flexibility and openness to the network, while has an impact on the inherent stability of the network. For SDN using, a overall consideration of the demand and balance in business systems for stability and flexibility should be taken into.

Key words: SDN; NovoNet; data center; SPTN

1 未来网络发展目标

通信网络经历了 1.0 模拟通信、2.0 数字通信、3.0 IP 化通信, 正式进入通信 4.0 时代。通信 4.0 需要信息技术(IT)和计算机技术(CT)融合。其核心特征包括敏捷化、开放化、软件化、虚拟化。为实现快速响应市场发展, 满足客户对通信网络深层次需求的目标, 亟须开展网络转型工作。中国移动于 2015 年初提出了发展下一代创新网络 NovoNet 的理念, 深度融合 IT 新技术, 构建“资源可全局调度, 能力可全面开放, 容量可弹性伸缩, 架构可灵活调整”的新一代运营商网络。

1.1 新型数据中心和新型网络

通信 4.0 时代的网络在物理设计上应具备两大基础特征:

(1) 网络功能软件化形成电信云, 构成新型数据中心。

(2) 软件定义网络(SDN), 实现灵活调度, 构成新型网络。

传统电信网络采用专用设备, 这些设备的电源、功耗、尺寸要求各不相同, 导致大量专用电信机房出现。未来的新型数据中心目标是统一规划, 采用通用服务器, 形成统一的资源池, 电信设备将采用网络功能虚拟化(NFV)技术实现软硬件分离, 电信网元功能以软件形式承载在统一资源池上形成电信云, 实现网络快速部署, 网元快速升级以及容量的弹性调度。新型数据中心的核心理念是 NFV 和 SDN。

传统 IP 网络自组织转发方式无法准确调度路由和开放路由; 轻载建设的传统模式在大带宽时代无法长久持续。新型网络采用 SDN 技术, 通过将路由由设备控制和转发功能分离, 实现网络路由的集中计算, 向转发设

备下达路由, 从而实现网络的灵活、智能调度, 以及网络能力的开放和可编程^[1]。

1.2 中国移动 NovoNet 计划

中国移动推出 NovoNet 新一代网络计划, 旨在融合 IT、CT 技术, 构建一张“资源可全局调度, 能力可全面开放, 容量可弹性伸缩, 架构可灵活调整”的新一代网络。NovoNet 中 Novo 是拉丁语创新的词根, Net 包括了移动和固定网络、IP 网络、传输网络以及数据中心网络。NovoNet 的核心技术是 SDN 和 NFV, 核心思想是“新架构、新运营、新服务”。NovoNet 将以新型数据中心为单元形成电信云, 通过新型网络实现高效承载, 最终实现网络功能软件化、资源共享

收稿时间: 2016-09-23
网络出版时间: 2016-10-31

化、网络可编程。

2 中国移动 SDN 部署实践

2.1 SDN 部署应用场景

中国移动 SDN 网络主要包括 5 个应用场景：

(1) NovoDC, 是 NovoNet 在数据中心的实现方案, 旨在为公有云、私有云提供多租户/多业务的虚拟网络, 满足用户自助开通虚拟私有云网络的需求。

(2) 软件定义的分组传送网 (SPTN), 是 NovoNet 在传送网的主要场景之一, 旨在为集团客户、长期演进 (LTE) 回传网络提供快速开通和质量保证。

(3) NovoWAN, 是 NovoNet 在广域网的实现方案, 通过实时感知网络流量, 全局集中调度流量, 提升 IP 网络带宽利用率, 并且能够保证关键业务质量。

(4) NovoVPN, 是 NovoNet 在集客多协议标签交换 (MPLS) 虚拟专用网络 (VPN) 的实现方案, 该方案基于 Overlay VPN 技术, 实现集客 VPN 用户的快速接入, 自动上线。

(5) 移动核心网、固网应用 SDN, 实现灵活业务链的编排。

2.2 数据中心 SDN 商用部署

2015 年中国移动公众服务云商用部署 SDN 网络, 主要提供私有网络 (VPC)、虚拟防火墙、虚拟负载均衡、VPN 等网络服务。该方案基于中国移动自研的 OpenStack 大云平台, 采用 Plug-in 方式和 SDN 对接。

该方案中 SDN 转发节点数百个, 控制器采用集群方式实现可靠保护和负载分担, 如图 1 所示。

2.3 SPTN 部署

2.3.1 SPTN 架构、北向接口和南向接口

SPTN 是业界率先提出面向超大

规模网络的层次化控制架构, 并付诸实现。SPTN 作为下一代城域传送网解决方案, 要能够覆盖全国所有县市, 实现多厂家, 百万节点组网。为了在大规模网络上, 同时实现区域内灵活管控和跨区域端到端调度, SPTN 引入了分层分域管理的层次化控制器模型, 如图 2 所示^[2]。

直接管理转发设备的控制器成为域控制器, 一个管理域对应一个域控制器来管控; 在域控制器之上, 可以继续构建超级控制器, 用于管理下层的多个域控制器; 若需要跨区域管理多个超级控制器的域, 还可以继续在此基础上构建更高层次的超级控制器。通过构建金字塔式的层次化控制器架构, 顶层控制器能对全网的集中控制中心, 实现大规模、多域、多厂家组网。同时, 各层控制器独立管控自己域内的资源, 能够实现本地运维和管理。

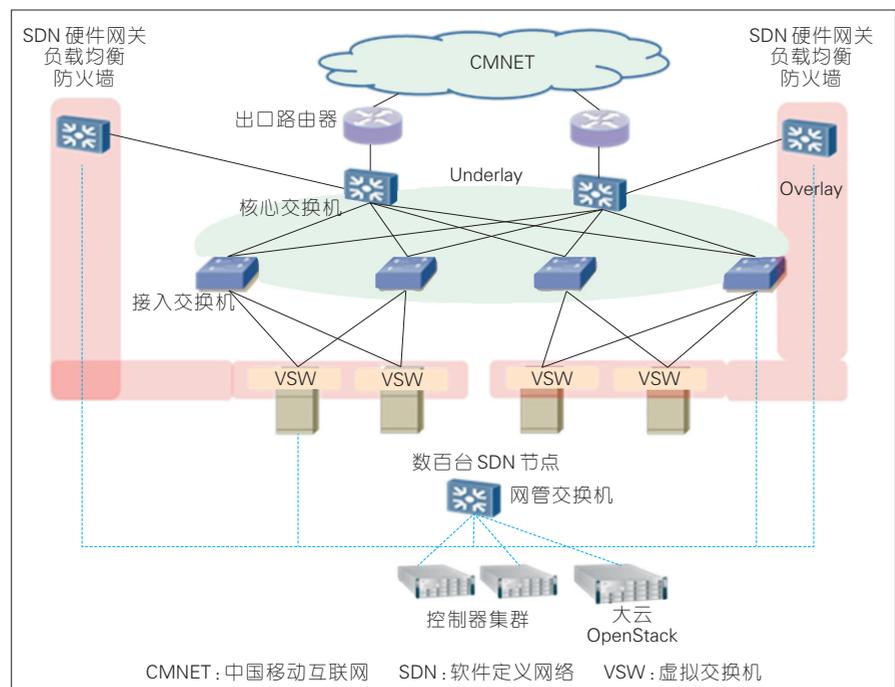
各层控制器采用统一的拓扑、资源、业务抽象模型, 上层控制器可通过下层控制器的北向接口的调用来完成自底向顶的集成。如果提升控制器的管理能力, 控制器支持管理更

多的转发设备的情况下, 可以避免控制器的层次过多, 这样更加有助于网络的集成和管理。

中国移动率先制订了 SPTN 标准信息模型及接口协议, 实现现网多厂家互联及智能化升级。北向接口指控制器之间以及控制器与应用程序 (APP) 之间的接口, 主要接口功能需包括拓扑管理、业务控制、告警、性能监测控制以及安全管理等功能。APP 与控制器之间的接口以及控制器之间接口的大部分功能相同, 为了减少接口标准的复杂度, 两个接口采用相同的标准协议。在 SPTN 系统中, 采用了业界主流的 RESTCONF 接口, 针对 SPTN 功能设计了统一的 YANG 数据模型, 描述接口的交互信息, 主要的模型如表 1 所示。

制订 SPTN 南向接口 (OpenFlow 表转发模型 (TTP)), 通过标准转发设备和南向接口实现跨厂家设备的直接管控。原生的 OpenFlow 无法满足 SPTN 电信级网络的要求。SPTN 针对 OpenFlow 做了如下方面的扩展：

(1) 转发 pipeline 扩展。该扩展支持网络侧接口 (NNI) 到用户网络



▲ 图 1 数据中心 SDN 网络架构

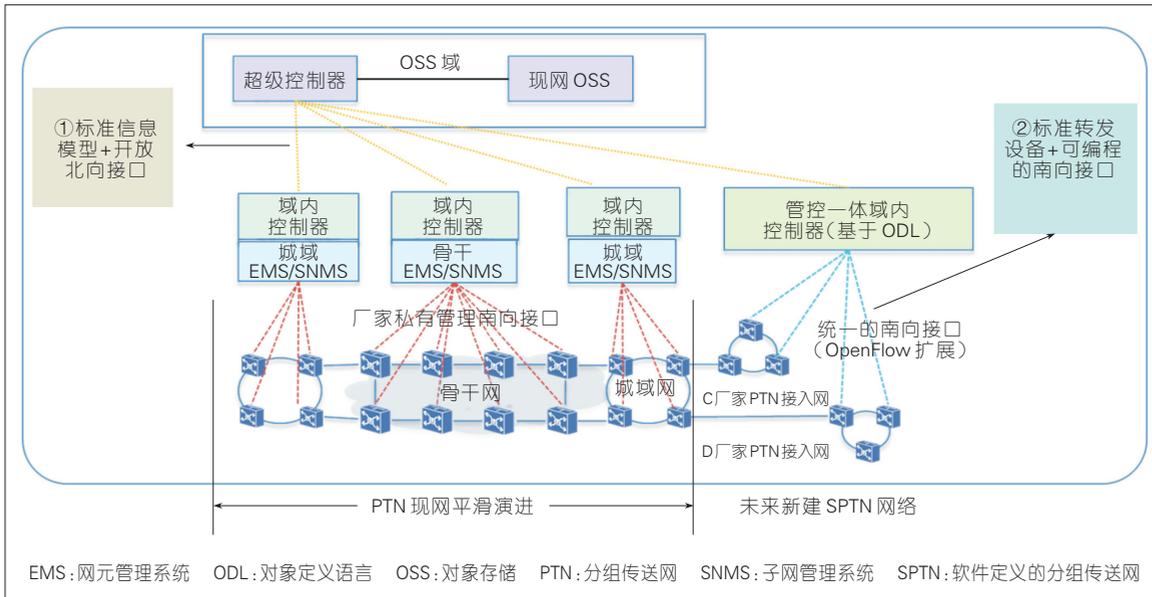


图2 SPTN 层次化控制器架构及标准接口

表 1 SPTN 北向接口主要 YANG 模型

属性对象	说明
隧道	两个端点之间一条隧道的行为及属性
以太服务流	某个以太业务的行为及属性
端点	某个 UNI 接口的行为及属
PW	两个端点之间一条伪线的行为及属性
APS 保护	某个 APS 实例的行为及属性
OAM 实例	某个 OAM 实例的行为及属性
NodeInfo	拓扑中某个节点的信息
内部 LinkInfo	拓扑中某个节点的控制器内部某条链路的信息
外部 LinkInfo	拓扑中某个节点的控制器外部某条链路的信息
APS: 自动保护倒换 OAM: 操作管理维护	PW: 伪线 SPTN: 软件定义的分组传送网 UNI: 用户网络接口

接口 (UNI) 侧的处理, 以及 UNI 到 NNI 的处理。这个处理过程能够使得 SPTN 系统能够插入或者提取合适的操作管理维护 (OAM) 报文或者保护报文。

(2) 对 OAM 的扩展。该扩展需满足层次化的 OAM 要求, 支持段层、伪线 (PW) 层、分层服务提供商 (LSP) 层以及业务层的各种 OAM 处理, 以及不同厂家设备之间的互联互通。对 OAM 的扩展模块应置于传送设备, 以硬件实现, 以保证实时性。

(3) 对线性保护倒换的扩展。线性保护需要运行较复杂的状态机, 如果都运行在控制平面, 50 ms 以内

的快速保护倒换很难实现。在 SPTN 中建议将倒换状态机的运行作为黑盒放置于传送设备, 该处理模块与 OAM 模块及 OpenFlow pipeline 对接, 同时, 控制器运行协同机制, 保证保护倒换状态的一致性。

2.3.2 SPTN 测试和试点

SPTN 构建了包括软件系统 (APP、超级控制器、域控制器)、硬件设备 (SPTN 设备及芯片), 测试仪表等的端到端生态系统, 如图 3 所示。

目前, 中国移动已经完成现网 7 厂家的实验室互通测试, 在广东、浙江、福建、北京等省份或地区进行了

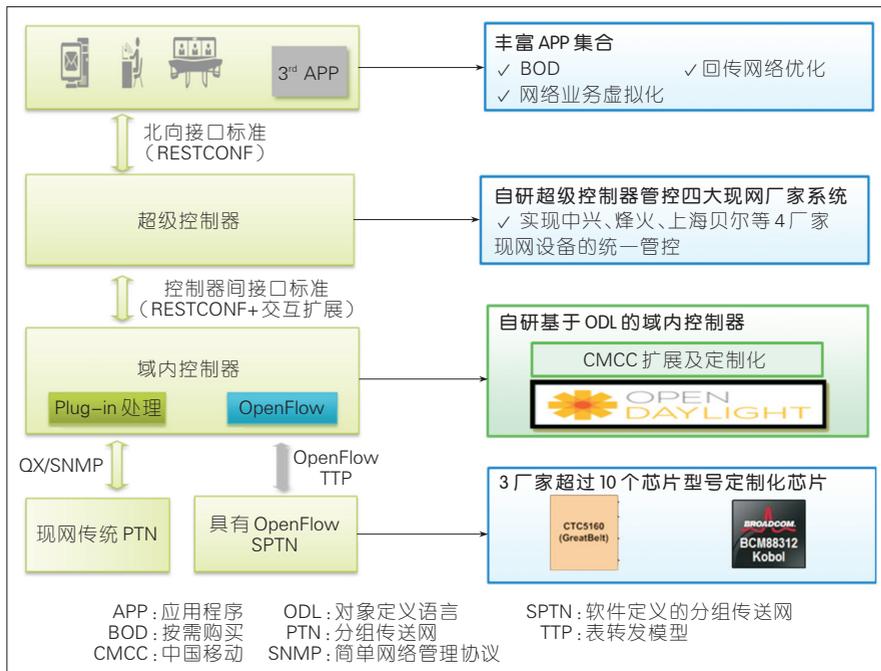
在现网商用试点。试点结果表明, SPTN 网络让 PTN 专线开通时间从“周”级别缩短到“分钟”级, 实现了跨厂家端到端业务开通。

2.4 自研 SDN 控制器

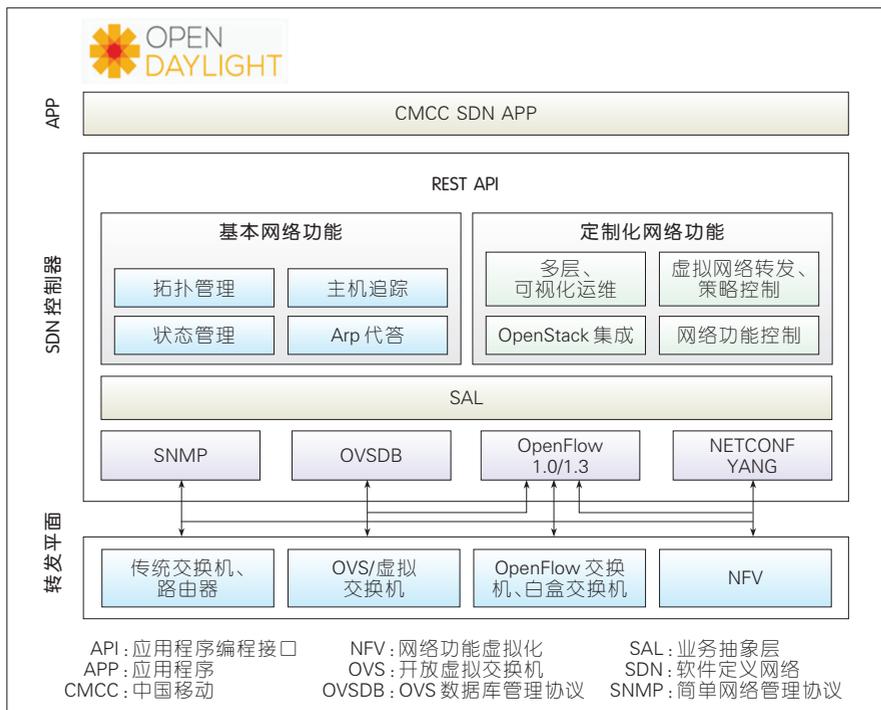
自 2014 年起, 中国移动基于开源 OpenDaylight 平台, 自主研发了面向数据中心的 SDN 控制器, 为公有云、私有云提供虚拟网络解决方案。自研 SDN 控制器一方面借助 OpenDaylight 平台实现多种南向协议 (包括 OpenFlow、NETCONF 等) 以及网络基本功能 (拓扑、状态管理、主机追踪等); 另一方面自主研发新功能, 满足中国移动数据中心的定制需求。自研新功能主要包括以下几部分, 如图 4 所示。

(1) 虚拟网络转发、策略控制。基于开放虚拟交换机 (OVS), 通过 OpenFlow 流表的方式实现功能包括 L2 转发、L3 分布式路由、安全组、虚拟端口限速、动态主机配置协议 (DHCP) 服务、元数据服务。

(2) 与第三方网元开放接口互通。利用多厂家商业 VXLAN 网关、防火墙、负载均衡器为虚拟网络提供服务。自研 SDN 控制器按照《中国移



▲图3 SPTN 架构和生态系统



▲图4 中国移动自研控制器架构

动SDN控制器接口规范》进行开发，控制商用设备。

(3) 多层、可视化运维。提供 Underlay 物理层、Overlay 虚拟网络层以及用户视图层3个层面的可视化

运维功能。

(4) OpenStack 集成。能够与主流 OpenStack 发布版 I、K 等集成。

中国移动自研 SDN 控制器不仅能够满足数据中心的定制化功能需

求,在性能方面也可以满足中型数据中心上千台虚拟机(VM)并发上线的能力。

3 SDN 规模部署关键问题

通过以上一定规模的商用部署和运营,以及真实环境下的压力测试,我们发现SDN网络在实际使用中还存在着一些问题亟待解决:

(1) 软件定义网络,除了把软件的优势带给了网络,也把软件的问题扩展到网络,影响到了网络的稳定性。数据库、消息队列以及异常处理是大型软件运行中影响稳定性的主要因素。传统网络设备采用专用转发芯片实现,操作系统运行的软件以及协议相对简单,稳定性较高。SDN网络中通过控制器集中处理消息,引入了消息队列、数据库等软件组件,在大压力下这些软件组件本身的问题将直接影响网络系统的稳定性。在使用中,我们发现了几个实际应用中的案例:

- 数据库不同步。OpenStack Neutron 和控制器都维护各自数据库。当一个虚拟交换机(VSM)故障,马上新建VM时,Neutron数据库来不及更新而控制器已经感知,后续对这个VM的处理将出现异常。

- 消息队列满。控制器消息队列满了,将造成后续VM端口启动消息丢失,进而导致后续VM无法执行DHCP流程并获取IP地址。

- 异常处理不完善。为了避免并发写数据时候的冲突问题,软件系统的操作方式有两种,一种是单数据库操作且写数据加锁,但这种方式效率有限;另一种方式是采用分布式数据库采用负载分担方式,不过容易出现并发写的时候锁不住,造成两边数据冲突,所以在锁不住的情况下往往采用回滚方式处理。由于控制器异常情况下回滚处理不完善,回滚失败,导致重新申请失败。

在大数据量并发配置的情况下,上述问题很容易发生,且直接影响网

络的稳定性。关于软件给网络稳定性带来风险的问题,需要重点研究和优化。

(2)目前的SDN产品不是开放,而是更封闭。传统网络中,基础网络基于以太网、IP路由协议搭建,主要采用多厂家数据中心交换机产品混合组网。对于安全、负载均衡等增值网络服务方面,可以通过IP技术兼容第三方厂家产品实现。从目前SDN产品看,包括控制器、虚拟交换机、SDN TOR交换机、SDN网关、防火墙、负载均衡、VPN设备等,都必须采用单厂家一揽子方案,最多强绑定指定第三方产品。这种格局并不利于SDN的长远发展。

该问题表现为2个层面的绑定:第1个层面是SDN和功能网元的绑定;第2个层面是SDN控制面和转发面的绑定。对于第1个层面,其本质在于目前缺少标准的业务链实现机制,所以厂家在SDN转发设备和防火墙、负载均衡等网络功能设备之间导流时,采用了SDN厂家私有技术和接口。另外,功能网元的配置定义不完善(如单臂双臂、子接口虚拟局域网配置等),需要相应的扩展。对于第2个层面,一方面从商业上讲,控制器和转发设备之间想打开是一种新的商业模式,还需要产业链自己去演进,也许未来会形成控制器厂家和转发设备厂家,而现阶段,每个厂家都同时有控制器和转发设备,想互相管理对方的设备难度很大;另一方面,单从技术上讲,即使不同SDN方案提供商宣称控制器和转发设备之间都支持OpenFlow1.3,但是在这些flow mod消息中到底相同功能的一条流表映射成为几级设备上的转发表,每级转发表匹配的内容和动作是什么,都是千差万别的,特别是控制器去配置硬件交换机时,还和芯片有关,难度更大。

中国移动自主开发SDN控制器的最主要需求就是为了解决上述两个问题。一方面通过定义标准的控

制器与多厂家转发设备之间的接口和模型,实现SDN和NFV的解耦;另一方面通过制订标准的OpenFlow流TTP,实现控制器与转发设备之间标准的南向接口,解耦SDN控制面和转发面。

(3)多域协同的架构选择问题。广域网部署SDN面临很多实际的挑战。电信级的网络中,网元数量众多,以PTN设备为例,每个省有几万端设备,全国近100万端设备,单一甚至集群控制器无法管理这么大规模的网络。另外,广域网的转发设备分布很广,覆盖全国31个省份,且网络拓扑复杂,省内分为接入、汇聚、核心3层,跨省还有单独的承载平面,如果控制器全国集中部署,光控制器和转发设备之间的传输时延就有几十毫秒,对控制面的可靠性和稳定性影响大。最后,由于长期网络的运维分工,广域网的运维、管控团队部门很多,较难实现理想的SDN集中控制和管理。

分域控制器是广域网引入SDN新技术,解决上述运营和维护问题的有效手段。分域控制器本身的方案选择也分为单层控制器+东西向接口、多层控制器+南北向接口两种方式。如图5所示,通过优劣势对比,

多层控制器平衡了复杂性,是有效的多域协同网络架构。

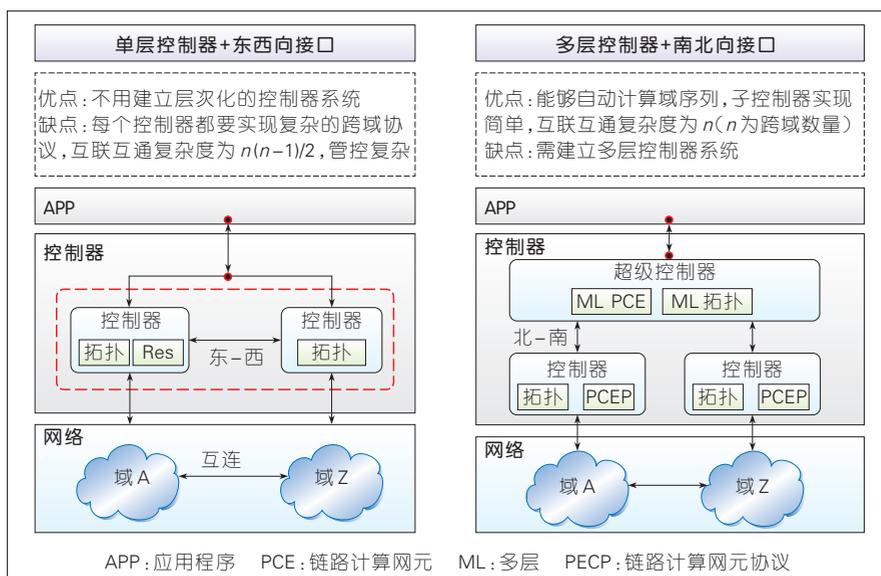
(4)缺少大规模复杂网络中控制器功能及性能测试方案。控制器的功能和性能测试是SDN方案的重中之重。在SDN产品快速发展的过程中,测试仪表和方案一直相对落后,没有专门的、通用的测试产品。为此,中国移动牵头梳理了SPTN场景下的测试需求,并联合仪表厂家、分组传送网(PTN)厂家联合开发SPTN测试仪表,填补了产业链在SPTN测试领域的空白。当前控制器测试的主要问题包括3点:

- 控制器是面向十万节点量级网络设计的,如何衡量控制器的性能指标并进行定量测试。
- 测试环境包括多层次化控制器及转发设备,如何进行全系统端到端精细测试。
- 各种接口是适配实际网络设计的,如何采用通用测试仪表进行相关测试。

测试仪表主要实现以下套件,完成接口、功能和性能测试。

- 接口一致性测试套件。分别模拟APP、各层控制器、转发设备,验证被测设备接口对象树构型、属性间

➡下转第39页



▲图5 单层、多层控制器方案对比