



云资源池网络自动化部署技术研究与实践

赖培源¹, 马卫民², 刘 艺¹, 陈 天¹

(1. 中国电信股份有限公司广州研究院 广州 510630; 2. 中国电信集团公司 北京 100032)

摘 要: 随着基于服务器虚拟化的云资源池规模的日益扩大, 网络成为资源快速交付的瓶颈点, 跨数据中心的业务互联互通及可迁移需求也日益突出, 云资源池的网络自动化部署技术成为新的研究热点。阐述了基于计算虚拟化技术承载的云资源池内网络自动化部署的概念、应用场景, 分析了基于软件定义网络理念的技术架构和关键技术, 总结了基于 overlay SDN 技术的解决方案和部署实践, 提出一种可运营的网络自动化部署架构, 并对电信运营商开展多业务、多租户承载的云资源池网络自动化部署技术选型给出了建议。

关键词: 云资源池; SDN; 多租户; 自动化; 电信运营商

doi: 10.11959/j.issn.1000-0801.2015185

Research and Practice on Network Auto-Deployment of Cloud Resource Pool

Lai Peiyuan¹, Ma Weimin², Liu Yi¹, Chen Tian¹

(1. Guangzhou Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China;

2. China Telecom Corporation, Beijing 100032, China)

Abstract: With the increasingly growth of cloud based resource pool, the network has become a bottleneck of the rapid service delivery. And the network automatic configuration technology has also become a new research topic as the requirement for inter service communication and migration across data centers becomes more imminent. The concept and applications scenarios of automatic network deployment based on server virtualization technology were described, and the key architecture and technology of software defined network were analyzed. Also the solution and practice of overly SDN technology were summarized, and an operational network architecture was proposed. Suggestion for multi-service provision and multi-tenant network of cloud resource pool automatic deployment was given.

Key words: cloud based resource pool, SDN, multi-tenant, automation, telecom operator

1 引言

现代社会信息量的爆炸式增长、资源复用技术的成熟和高带宽网络的普及, 共同促进了云计算的诞生和发展^[1]。

当前, 随着 x86 服务器虚拟化技术不断成熟、商业模式获得广泛认同, 硬件重构和软件定义技术正引发资源部署方式和运营能力的巨大变化。云资源池规模快速扩大, 在虚拟机形态的计算、存储资源快速交付的同时, 租户虚拟机

收稿日期: 2015-05-20; 修回日期: 2015-07-13

论文引用格式: 赖培源, 马卫民, 刘艺等. 云资源池网络自动化部署技术研究与实践. 电信科学, 2015185

Lai P Y, Ma W M, Liu Y, et al. Research and practice on network auto-deployment of cloud resource pool. Telecommunications Science, 2015185

组网及安全策略等配置仍需由云资源池管理员进行配置,历经需求沟通、手工配置、确认、更新等多个环节,网络配置周期长,且对云资源池运维人员提出了较高的综合技术水平要求,网络成为 IT 资源快速交付和规模运营的瓶颈点。因此,以网络虚拟化为代表的云数据中心网络自动化部署技术成为新的研究热点。

现有的网络虚拟化方法,如虚拟局域网(virtual local area network, VLAN)、虚拟路由转发(virtual routing forwarding, VRF)表等,受协议自身的限制(如 VLAN 数量)、设备容量限制(MAC 表容量)、管理的兼容性、开放性以及自动化能力不足等方面的影响,无法满足大型云数据中心大量虚拟机间动态组网及多租户承载需求。软件定义网络(software defined networking, SDN)是为了适应云计算在计算虚拟化后对网络虚拟化的需求而发展起来的,其控制转发分离、逻辑集中控制的网络架构,天然适合实现网络资源的抽象、池化和自动化部署。SDN 通过软件方式在物理网络上构建虚拟的逻辑网络,使得上层业务对网络的配置、变更需求直接体现在逻辑网络上,屏蔽了业务对底层物理网络的影响。

然而,作为一项新兴的技术,SDN 存在以 OpenFlow^[2]、overlay^[3]和 I2RS^[4](interface to the routing system)等为代表的多种技术流派,且皆有较为完整的技术方案。在云资源池建设中,如何在尽可能保护既有投资、降低网络改造成本和风险的前提下进行技术选型,是一个亟需研究的问题。本文首先阐述了基于计算虚拟化技术承载的云资源池内网络面临的问题和自动化部署的概念及相关技术,并分析了基于软件定义网络理念的技术架构和关键技术,总结了基于 overlay SDN 技术的解决方案和部署实践,最后提出一种可运营的可靠性高的 overlay SDN 部署架构,并对电信运营商开展多业务、多租户承载的云资源池网络自动化部署架构和技术选型给出了建议。

2 云资源池网络部署面临的问题

随着计算虚拟化技术的成熟和规模化商用,经过前几年的业务平台云化承载验证,在企业私有云环境下承载的业务系统已有了较大的规模,基本形成了较为完善的业务部署流程。然而,随着资源池规模的扩大,承载业务平台数量剧增,云资源池网络部署面临着诸多问题。

(1) 交付效率低

资源池网络仍采用接入、汇聚/核心的二级/三级架构,

租户网络的配置通常需要同时在云管理平台和物理网络设备上配置,由于资源池内部网络设备众多,包括不同厂商的交换机设备、防火墙设备、负载均衡设备等,加之业务属性复杂,网络需求各不相同,不同业务系统的流量、安全等策略不同,基于繁杂的点对点手工配置,严重拖累业务上线进度,且对系统排障带来很大的挑战。

(2) 运维难度大

资源池网络配置需要由云资源池运维人员进行统一管理,而由于缺乏总体的拓扑展现,业务系统与网络资源无法呈现对应关系,运维人员大多不熟悉具体的业务系统,导致分工不明确,大幅增加了沟通成本,运维复杂,难以满足用户灵活部署和调整网络的诉求。虽然一些网络虚拟化技术可以支持网络的多租户能力,但其权限开放程度和系统性能都面临新的问题,如租户网络的南北向流量和安全策略开放可能会影响到其他租户甚至整个数据中心的网络安全和稳定性,而基于虚拟机的网关形态,在架构上造成了流量的绕行且成为了资源的瓶颈点。

(3) 灵活性不足

云资源池以“按需使用、错峰填谷”的弹性特征获得广泛认可,其承载的业务类型众多,网络负荷同样存在灵活调度的需求,而由于资源池网络无法动态感知业务属性或与业务负荷形成协同管理,造成资源浪费或过载。

(4) 隔离规模小

云资源池网络一般使用 VLAN 进行租户网络的隔离,这在规模较大的云资源池环境下会有两大限制:一是由于 VLAN 数量在标准定义中只有 12 bit,即 VLAN 的最大数量被限制在 4 096 个以内,随着虚拟机数量的快速增长,加之业务网、管理网、存储网等多网络的隔离需求,这个数量对于公有云或规模稍大的企业云计算应用而言远远不足;其次是 VLAN 技术当前为静态配置型技术,虽然通过虚拟交换机或虚拟机接入感知(VEPA/IEEE 802.1Qbg)技术也可以支持 VLAN 的动态配置,但通常需要将物理交换机上的端口对应更新配置,因为如果配置为所有 VLAN 通过,则会造成大量的广播数据和安全隐患。

此外,作为云计算的基础,虚拟化的一大优势就是虚拟机的在线迁移,虚拟机可以从一台物理主机迁移到另一台物理主机上运行,保持业务不中断。而当前的云资源池网络架构大多只能支持同一个数据中心二层网络内的迁移,对于混合云、云灾备等需要进行整个租户 IT 环境整体迁移的应用场景,则难以满足业务需求。由此可以看出,云



资源池要求网络具备快速部署能力、支持多租户的自服务、易管理能力,同时支持底层物理网络的松耦合,支持网络的灵活调配、弹性伸缩,甚至希望云管理平台可以实现资源池内虚拟交换机、物理交换机、防火墙以及负载均衡器等网络设备的统一管理,包括资源配置表单的统一配置以及相应配置命令的下发,最终实现全部网络设置在一个界面上完成自动化的高效配置,为进一步实现智能化自动管理提供实施基础和统一的管理能力。

3 网络自动化部署技术

云资源池网络自动化部署主要指基于业务的需求,由系统自动实现逻辑网络资源与物理网络资源的转化,通过配置表单、逻辑拓扑等可编排模式自动实现业务系统的网络部署。如图1所示,用户只需对虚拟机的虚拟网卡(virtual dual NIC, vNIC)进行表单配置,包括IP地址、端口、安全策略、负载均衡策略等,并对多台虚拟机之间的隔离组进行可编辑模式下的拓扑配置,即可完成业务系统的网络部署。

当前,云资源池运维部门通常以业务系统的维度为租户/用户组建一个虚拟的逻辑网络,从而为用户提供一个独立的IT环境,即用户能基于创建的虚拟机进行灵活的组网,并制定相应的网络策略,就像是在传统数据中心内使用交换机对服务器进行组网一样。在传统的网络环境下,当前使用的网络自动化部署技术主要有以下几种。

- 利用VLAN进行租户的网络隔离,使用虚拟机形态的网关进行组网管理,如CloudStack高级网络模式中的系统虚拟机vRouter^[5]。在CloudStack的资源池中,几乎所有虚拟网络功能都是vRouter实现的,如DHCP、DNS、NAT、防火墙、端口转发、负载均衡、VPN等。每个租户网络至少对应一个vRouter,当启动某个租户网络的第一台虚拟机时,vRouter会自动创建出来,并作为虚拟网络的网关,通过在CloudStack上对其进行配置,即可完成网络的自动化部署。这种技术可以认为是网络功能虚拟化^[6](network function

virtualization, NFV)的一种实现,但由于系统虚拟机的性能、VLAN数量等多方面的限制,目前还难以满足大规模云数据中心的部署要求。

- 通过云管理平台对云资源池的网络设备进行纳管,定制化开发设备管理系统,并通过API调用或自动化配置脚本的方式完成对物理网络设备的统一管理。这种技术可以认为是I2RS的一种实现。但由于各云资源池的网络设备大不相同,针对每一款设备进行定制开发不仅需要投入大量的开发费用,还需获得设备厂商的技术支持和接口开放,且存在设备纳管的安全性问题,一旦脚本命令稍有差错,则可能导致整个网络的中断和不稳定。以OpenStack为代表的云管理平台以其开放架构为基础,可以通过支持设备厂商提供管理平台插件的方式实现对设备的纳管,但目前还没有相关实现。
- 采用overlay SDN技术,在现有网络架构上叠加逻辑网络,将物理网络池转化成逻辑资源,租户可以在一个由隧道构成的逻辑网络内按需配置、管理自己的业务网络,同时该逻辑网络也可以与计算资源、存储资源一起被云管理平台统一调度和按需交付,实现多租户环境下灵活的网络资源提供。

虽然从技术架构上看,OpenFlow SDN也能解决现有云资源池的网络问题,但需要新购支持OpenFlow协议的物理设备,对现网的改动要求较大,且暂无成熟的商用解决方案,存在较大的风险,因此,下文重点介绍overlay SDN技术和相关解决方案。

4 云资源池 overlay SDN 架构及关键技术

overlay SDN通过软件的方式在现有网络架构上叠加逻辑网络,采用隧道方式对转发面进行抽象,对基础网络不进行大规模的修改,实现业务与承载分离。转发面被抽象为边缘层和核心层,边缘层(隧道终节点)负责业务接入及网络服务提供,核心层负责高效、低时延的数据传输。控制面上引

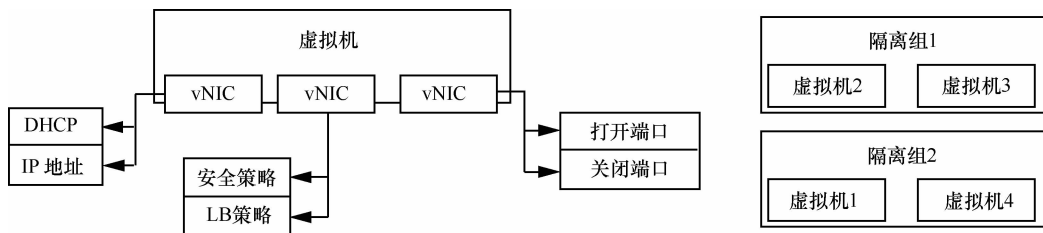


图1 云资源池网络自动化部署模型

入集中控制器,控制器向下通过南向协议对逻辑网络进行控制,向上通过北向接口开放逻辑网络的编程能力。

4.1 云资源池 overlay SDN 逻辑架构

overlay SDN 是在虚拟化的服务器上引入了一个 TEP (tunnel endpoint,隧道终端)的逻辑体,完成隧道报文的封装和解封装。如图 2 所示,TEP 与物理网络相连,且分配有物理网络的 IP 地址,该地址与虚拟网络无关,多个 TEP 可以为租户组建一个架构在隧道上的虚拟网络。当逻辑网络内的两台虚拟机要进行通信时,虚拟机的数据分组在 TEP 上被封装为以 TEP 的 IP 地址为源地址的数据分组,加上外层的 MAC 地址头、IP 地址头、TCP/UDP 头进行发送,物理网络按照正常的路由进行数据分组的转发,数据分组到达目的 TEP 后,TEP 进行数据分组的解封装并把数据分组发送给目标虚拟机,完成一次通信过程。其中,源 TEP 和目标 TEP 地址的获取和管理均可由集中的网络控制器来实现。为了让 overlay 虚拟网络之间以及虚拟网络与物理网络之间能够进行通信,通常网络中还会有一个 overlay 网关,用于终结隧道报文,实现虚拟网络和外部网络(非 overlay 网络)的互通。

TEP 的功能可以是服务器虚拟化软件中的一个功能模块或一台虚拟机,也可以在物理网卡或者物理交换机上实现。在 overlay SDN 解决方案中,隧道的封装协议、控制器的功能和性能、TEP 的性能和非 overlay 网关能力是整个网络能力的关键因素。

4.2 隧道封装协议

在隧道封装协议选择上需要考虑的因素有很多。在技术层面上,要考虑新增数据分组头的开销,它影响着网络的传输效率和成本;也要考虑隔离标识的长度,它决定了可支持的隔离规模;此外,还要考虑是否需要改动物理设备或对物理设备有特殊的要求(如支持 GRE 等)。非技术

层面的因素需要更多的关注,尤其是标准的行业支持程度,它涉及标准协议的生命力和产品的兼容性等。

目前 overlay SDN 采用的主流隧道协议主要有虚拟可扩展局域网(virtual extensible local area network,VxLAN)、无状态传输隧道(stateless transport tunneling,STT)、通用路由封装的网络虚拟化(network virtualization using generic routing encapsulation,NVGRE)、Geneve (generic network virtualization)4 种,见表 1,4 种协议之间有很多相似之处。VxLAN 是将以太网报文封装在 UDP 传输层上的一种隧道转发模式,NVGRE 采用通用路由封装(generic routing encapsulation,GRE)协议,将以太网报文封装在 GRE 内进行隧道传输。两者的虚拟网络 ID 均为 24 bit,可以支持大约 2^{24} 个虚拟网络。STT 是无状态传输协议,通过将以太网报文封装成 TCP 报文进行隧道传输,隔离标识采用 64 bit 的字段,可以支持大约 2^{64} 个虚拟网络。Geneve 是由 VMware、Microsoft(微软)、Redhat 和 Intel(英特尔)联合提出的 overlay 协议,旨在综合现有的主流 overlay 协议(VxLAN、STT 和 NVGRE)的优势,Geneve 采用 UDP 封装方式,草案提出的虚拟网络标识长度为 24 bit。

目前的 VxLAN 技术尚未明确定义控制平面,在缺乏控制平面的情况下,地址学习需要 VTEP 自主进行。由于 VTEP 需要通过多播在 VTEP 之间进行地址学习,而大部分网络设备未开启多播功能,且多播功能影响网络设备性能、存在安全问题等,因此受限较大。各大厂商在推出自己的 VxLAN 产品时,也在着力解决多播问题,大部分厂商提出增加一个控制器,实现 VxLAN 控制层面集中管理,由控制器汇总 VTEP 信息,下发转发路由到 VTEP,在这种情况下,无需通过 VTEP 间多播确定转发路由。目前,支持 VxLAN 的产品有 VMware 的 NSX、H3C 的 VCF、ALU/Nuage 的 VSP、华为的 FusionNetwork、Cisco 的 Nexus 1000v、ACI、

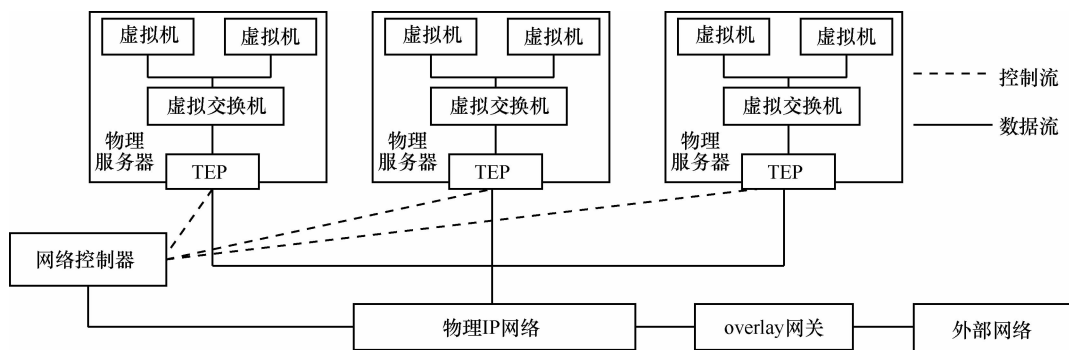


图 2 overlay SDN 网络模型



表 1 4 种 overlay 隧道封装协议对比

技术名称	支持者	标准状态	新增报头长度 / byte	虚拟网络标识长度/bit	技术本质	链路散列能力
VxLAN	VMware、Cisco(思科)等	IETF 标准 RFC7348	50(+4 可选)	24	L2 over UDP	现有网络可以运行 L2~L4 散列
STT	VMware(Nicira)	在 IETF 已过期	58~76(+4 可选)	64	L2 over TCP(无状态)	现有网络可以运行 L2~L4 散列
NVGRE	Microsoft、Arista、Intel、Dell(戴尔)、HP、Broadcom(博通)和 Emulex 等	IETF 草案中	42(+4 可选)	24	L2 over GRE	GRE 头的散列需要网络升级
Geneve	VMware、Microsoft、Redhat、Intel 等	IETF 草案中	54(+4 可选)	24	L2 over UDP	现有网络可以运行 L2~L4 散列

Arista、博科的交换机以及中兴通讯的虚拟化软件 ZXCLoud iECS 等。

STT 的主要提出者是 Nicira,随着 Nicira 被 VMware 收购,整体转向支持 VxLAN 封装,目前 STT 标准在 IETF 已过期,且日益淡出视野。NVGRE 是一个使用了 GRE 协议封装在三层网络上隧道传输二层网络分组的解决方案,其主要提出者是微软,支持厂商包括 Arista、Intel、Dell、HP、Broadcom 和 Emulex 等,目前已在 IETF 草案中。随着 VxLAN 协议正式成为 IETF 标准,NVGRE 前景看淡。Geneve 的目标是成为一个纯粹的隧道格式规范,它与其他 overlay 协议一样,定位于解决在数据中心内网络可扩展性、租户隔离、虚拟化感知、逻辑网络与物理网络解耦等问题,同时又考虑到未来功能的扩展性,可用于数据中心内互联、数据中心间互联等应用场景。虽然目前尚未有相关商用产品支持 Geneve 协议,但其主要提出者中,VMware 和 Microsoft 也是三大主流 overlay 的提出者(VMware 是 VxLAN 的主要提出者,Microsoft 是 NVGRE 的主要提出者,而 STT 的主要提出者 Nicira 已被 VMware 收购),Geneve 的前途还是比较乐观的。

4.3 控制器的功能和性能

overlay 技术尚未明确定义控制平面,而 SDN 控制器的集中管理模式能有效解决 overlay 中的多播问题,同时还能通过可编程等特性提升网络的灵活性。控制器是 SDN 架构的核心,是一个逻辑上集中的设备,其作用类似于计算机操作系统,用于管理和控制底层网络资源,包括物理网络设备和软件形态的网络设备,同时为应用程序提供使用网络资源的接口,包括上层的云管理平台或网络编排系统等,并维护全网拓扑,根据业务需求计算下发流表,并提供相关网络服务。

如图 3 所示,SDN 控制器功能通常可以分为 5 个层

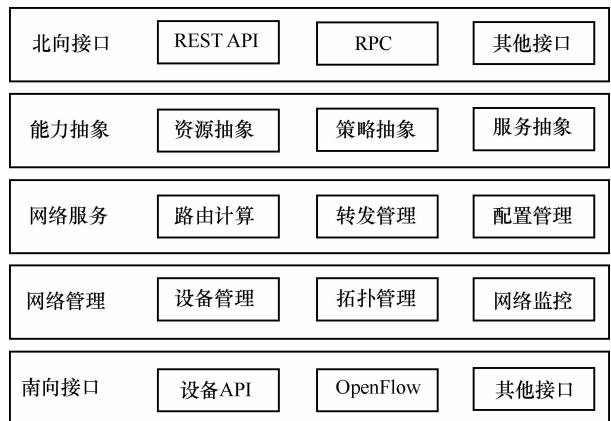


图 3 overlay SDN 控制器功能架构

次,具体介绍如下。

(1)南向接口

封装控制器与底层网络设备交互的接口、协议,为控制器其他模块访问网络设备提供统一的接口,常见的南向接口有 OpenFlow、XMPP、OVS-DB、NETConf、OF-Config 等。控制器可采用插件(plugin)技术实现对多种南向接口协议的扩展支持,通常每个插件封装一种协议。南向接口能力决定了控制器可纳管的设备范围和功能以及系统的可扩展性。

(2)网络管理

通过发现服务获取底层网络的交换机、链路和主机信息,构建全局的网络拓扑,为控制器的其他模块提供全局的网络信息,并负责网络的监控,包括设备、链路的运行状态、日志和事件记录等,从而保持网络信息的时效性。

(3)网络服务

基于全局网络拓扑和实时动态的网络状态信息,控制器可以自主或按照应用程序的要求灵活定义转发路径,并选择最优的转发路径,同时负责和转发设备的交互模式,按需下发或更新数据分组的转发路径,调整网络策略。

(4)能力抽象

根据业务定义自动构造用户逻辑网络,将各种逻辑网络服务单元进行有序的安排和组织,实现逻辑网络和物理网络的解耦。根据业务需求,控制器需要按照一定的算法、策略完成资源的编排,高效提供各种逻辑网络,资源的组合及负载均衡能力很大程度上影响着网络的整体性能。

(5)北向接口

通过标准接口为应用程序开放网络资源、策略和服务,目前业界尚未形成北向接口标准,不同控制器对外开放的能力不同,开放的方式也不同,主要取决于控制器的实现方式。RESTful 接口因其易用性和跨平台的兼容特性被广泛采用。北向接口的标准化影响着在大规模云数据中心中能否统一纳管多个 SDN 解决方案以及业务系统能否在不同 SDN 域进行迁移。

由于单台控制器能够支持的网络规模有限,实际部署中大多数是多台控制器组成的控制器集群,按区域协同管理和控制底层网络,从而提升网络的性能、可靠性和可扩展性。可见,控制器集群间的协同能力、负载均衡能力、主备切换能力也是选择控制器的主要考虑因素。

4.4 TEP 的兼容性和性能

作为 overlay SDN 的重要组件,直接和虚拟机连接的 TEP 组件的兼容性和性能无疑也是重要的影响因素。TEP 组件的兼容性通常包括服务器虚拟化软件的兼容性、提供给控制器的接口兼容性以及不同 TEP 间通信的兼容性 3 个方面。软件 TEP 目前主要有两种部署形态,一种是代替虚拟交换机直接嵌入物理服务器的虚拟化层,通常称为内核态;另一种是以虚拟机的形态部署在物理服务器的虚拟化软件之上,同一服务器上的其他虚拟机通过将流量绕行到 TEP 上进行封装管理。从技术架构上来看,内核态的 TEP 通常能获得较好的性能,但其兼容性受到虚拟化软件厂商的支持限制。而虚拟机形态的 TEP 由于流量的绕行和虚拟机的转发能力限制,交换性能较差,但在兼容性上受到虚拟化软件厂商的影响较小,然而大多情况下仍需要根据不同的虚拟化软件开发对应的 TEP 虚拟机,从而获得较好的性能。相比软件 TEP 通过服务器 CPU 能力完成原始以太报文的封装和解封装,硬件 TEP 的性能和兼容性会更好一些。

当数据中心网络需要引入多个厂商的 overlay SDN 方案时,则需要明确 overlay SDN 方案与其他厂商 SDN 方案的兼容能力。由于不同厂商在南向接口的实现上仍存在较大的差异,即便是使用同一版本 OpenFlow 协议的解决方案,

仍无法实现标准化,不同厂商的 SDN 控制层间尚不存在可用于对接的标准协议,由此 VTEP 的信息也无法传递到其他厂商的 SDN 控制层。因此,目前的互通方案只能将各不同厂商的 overlay 网络各自终结在一个管理域内,将非自身管理域内的网络视作外部物理网络,通过网关实现互通。

4.5 overlay 网关

overlay 网关主要作为南北向流量的出口,实现 overlay 网络与外部网络(非 overlay 网络或不同厂商的 overlay 网络)的互通。当外部网络帧准备进入 overlay 网络时,overlay 网关将该帧进行封装,转换为 overlay 帧转发;当 overlay 帧准备进入外部网络时,VxLAN 网关将该帧进行解封装,转换为非 overlay 帧转发。与 VTEP 类似,overlay 网关同样可以以软件或者硬件的方式部署,可以是承载在通用服务器上的软件,也可以是接入交换机、汇聚交换机或者更高层的交换机甚至边缘路由器。从目前的实现来看,硬件网关的性能要高于软件网关,但软件网关的成本较低,部署灵活,在小规模应用场景下仍有一定的优势。

overlay SDN 以 IP 网络为基础,通过成熟的隧道技术叠加逻辑网络,对底层承载网络无要求,主流厂商的 SDN 商用产品主要集中在这个方向,其商品成熟度最高。虽然业界已有不少 overlay SDN 商业解决方案,但这些方案在功能、性能、开放性和兼容性等方面都还有缺陷,对虚拟化平台和云管理平台的适配能力也还不足。

5 overlay SDN 技术发展和产业生态

目前 overlay SDN 方案已有若干商用解决方案,且已在国内外运营商和互联网企业的云资源池和云数据中心进行部署验证。虽然各个厂商的解决方案总体架构差异较小,但具体实现上却有较大差异,且具有一定的技术发展和产业格局代表性,下面介绍几种主要的商用解决方案。

(1)VMware 的 NSX 解决方案

目前的最新版本是 2014 年 9 月发布的 NSX for vSphere 6.1,主要组件包括管理器、控制器、网关和 vSwitch。NSX 管理器在网络虚拟化环境中为 NSX 提供单个配置点和 REST API。控制器可对数据平面的虚拟交换机的二层通信以及东西向路由的三层通信进行控制。NSX 网关是基于 x86 架构、以虚拟机方式部署的三层网关,用于实现虚拟网络和物理网络的连接。NSX 中的 vSwitch 是基于分布式交换机(vSphere distributed switch,VDS)结合其他组件实现,利用 VxLAN 协议支持隧道网络,结合其他



组件提供分布式路由、分布式防火墙等服务。

(2) ALU/Nuage 的 VSP 解决方案

目前最新版本是 2014 年 7 月发布的 R2.1, VSP 主要由虚拟网络服务目录(VSD)、虚拟网络服务控制层(VSC)、虚拟网络路由交换机(VRS)和网关 4 部分组成。VSD 是一个网络策略引擎,主要向用户提供基于 Web 的管理界面,用于控制网络的配置,并通过 XMPP 向 VSC 下发策略;VSC 是 SDN 控制器,负责通过 OpenFlow 协议向 VRS 下发流表;VRS 是基于 Open vSwitch 实现的虚拟交换机,支持 KVM 和 Xen 的内核态部署以及 ESXi 环境下的虚拟机形态部署。此外,VSP 解决方案还提供专用硬件网关 VSG 和基于 x86 架构实现的软件网关 VRS-G,用于虚拟网络和物理网络的连接。

(3) 华为 FusionNetwork

主要包括管理器 FNM、控制器 FNC、虚拟交换机 EVS、NFV 节点和网关等组件。FNM 是基于 Neutron 实现的,负责 2~7 层的业务协同,并提供标准的 API;FNC 是 SDN 控制器集群,通过 OpenFlow 协议向 EVS 下发流表,实现 2~4 层的网络业务能力;EVS 是基于 Open vSwitch 实现的增强虚拟交换机;NFV 节点支持各类 NFV 业务,如防火墙、负载均衡、VPN 等;网关既支持虚拟二三层网关,也支持物理二三层网关。

(4) H3C 的 VCF

主要架构组件包括 iMC、VCFC、H3C vSwitch 和网关。其中,iMC 负责向用户提供基于 Web 的上层管理界面,并提供业务链服务管理;VCFC 控制器主要负责向上提供 REST API,通过 OpenFlow 协议下发流表;H3C vSwitch 是基于 Open vSwitch 实现的增强虚拟交换机;网关支持软件和硬件两种形态,软件网关 VSR 是基于 x86 架构实现的软件三层网关,用于虚拟网络和物理网络以及虚拟网络内的跨三层通信,硬件网关分两种,二层网关用于 VxLAN 与 VLAN 环境连接,三层网关用于 VxLAN 与传统网络跨三层互联。

当前的几种 overlay SDN 商用解决方案都基本具备了管理器、控制器、TEP 和网关等基本组件功能,在中国电信的私有云资源池中,经过在数百台虚拟机规模的试验环境验证,单厂商的 overlay SDN 解决方案能有效提升云资源池的自动化水平和资源的交付效率,可提供更灵活的业务隔离、流量监控等网络安全管控,实现更直观、统一、高效的网络资源管理。然而,不同厂商的 SDN 网元之间仍基本无法互通,服务器虚拟化软件的接口开放程度仍制约着 TEP

的发展,overlay 软件网关的性能和功能还有待进一步提升,基于软件定义的开放式部署架构仍有很长的路要走。

6 电信运营商网络自动化部署架构设计和策略

虚拟化技术改变了 IT 基础设施的提供模式,使得计算、存储设备变成可运营的资源,用户可以通过互联网按需弹性获取。基于 SDN 技术可构建完全虚拟化的网络并实现网络资源的自动化编排,并能够带来一系列资源重构的效益,如低硬件投资成本、智能化基础设施、提升资源交付效率等,然而对着重于整合运营、自主开发能力有限的电信运营商来讲,也存在着诸多挑战,如软件层面可能形成新的垄断格局、软硬件成本叠加后的整体成本未必下降等。可见,接口、协议的标准化是规模化运营、真正实现架构开放的重要环节。

实践证明,单厂商的 overlay SDN 商用解决方案能有效缓解当前云资源池网络面临的问题,满足业务和运维需求,因此,在现有条件下可以采用如图 4 所示的部署架构,在单一集群中部署单厂商的 SDN 解决方案,并通过在云管理平台中部署管理插件的方式实现集中管理,不同 SDN 厂商解决方案的网络间通过网关进行通信。

插件管理模式也非长久之计,随着插件类型的增多,管理平台的负荷将会不断增大,且可能存在一定的安全问题,也无法满足扩容方案的迁移。因此,长期来看,电信运营商还需积极参与推动 overlay SDN 的标准化发展,在厂商的解决方案选型上,重点考虑技术成熟度、发展趋势、对现有网络的改动、兼容性要求、接口标准化程度、统一管理方式、使用成本、各模块功能及性能等问题。此外,还需跟踪 OpenFlow、NFV 等技术的发展,加大自主研发投入。

7 结束语

随着物联网、移动互联网技术和业务的快速发展,IT 资源需求呈爆炸性增长,IT 基础设施服务在规模、运维等方面面临着诸多挑战。在云计算逐步落地之后,除了虚拟化较为成熟之外,存储、网络等其他资源的池化仍有一定差距,尤其是网络自动化部署需求日益强烈,虽然目前基于 overlay 的 SDN 解决方案能够带来一定的便捷,但在管理层面(如网络能力的抽象、服务的编排、网络模板等)、控制层面(如集中控制的拓扑管理、路由算法等)、转发层面(如网关能力、虚拟交换机性能等)以及接口的标准

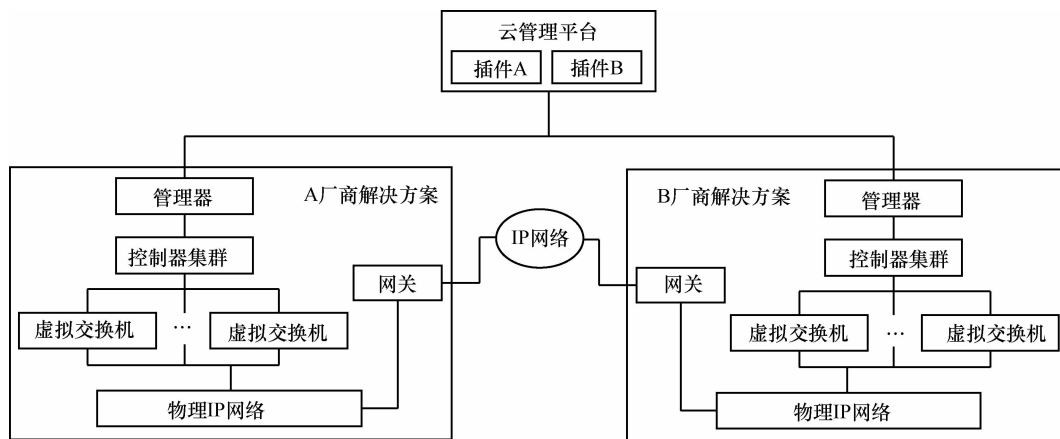


图 4 电信运营商网络自动化部署架构设计

化层面,均处于起步阶段。同时,随着计算虚拟化的成熟和广泛应用,跨数据中心的业务互联互通及可迁移需求也日益突出。推动计算、存储、网络、安全等资源的综合智能调度,诸如人工智能、大数据分析、流量工程等技术研究也在持续进行中。可以预期,在各种业务需求的推动下,随着相关技术的发展与成熟,整个基础设施环境将全面走向自动化,实现资源到服务的交付。

参考文献

- 1 李丹, 刘方明, 郭得科等. 软件定义的云数据中心网络基础理论与关键技术. 电信科学, 2014, 30(6): 48~59
Li D, Liu F M, Guo D K, *et al.* Fundamental theory and key technology of software defined cloud data center network. Telecommunications Science, 2014, 30(6): 48~59
- 2 McKeown N, Anderson T, Balakrishnan H, *et al.* OpenFlow: enabling innovation in campus networks. ACM Sigcomm Computer Communication Review, 2008, 38(2):69~74
- 3 Stephens B, Cox A, Felter W, *et al.* PAST: scalable ethernet for data centers. Proceedings of ACM Sigcomm Conext Conference, Nice, France, 2012
- 4 Hares S, White R. Software-defined networks and the interface to the routing System (I2RS). IEEE Internet Computing, 2013, 17(4):84~88
- 5 Cloudstack 中国社区. 深入系统虚拟机之 vRouter. <http://www.cloudstack-china.org/2012/09/650.html>, 2012
Cloudstack Chinese Community. vRouter of thorough system virtual machine. <http://www.cloudstack-china.org/2012/09/650.html>, 2012
- 6 ETSI. Network Functions Virtualisation-Introductory White Paper, 2012
- 7 ONF. OpenFlow Switch Specification 1.3.1, 2012
- 8 Hassas S, Ganjali Y. Kandoo: a framework for efficient and

scalable offloading of control applications. Proceedings of First Workshop on Hot Topics in Software Defined Networks, Helsinki, Finland, 2012: 19~24

- 9 左青云, 陈鸣, 赵广松等. 基于 OpenFlow 的 SDN 技术研究. 软件学报, 2013, 24(5): 1078~1097
Zuo Q Y, Chen M, Zhao G S, *et al.* Research on OpenFlow-based SDN technologies. Journal of Software, 2013, 24(5): 1078~1097
- 10 樊勇兵, 丁圣勇, 陈楠. 基于业务交换机的大规模云数据中心通用网络架构设计. 电信科学, 2013, 29(10): 1~4
Fan Y B, Ding S Y, Chen N. A general network architecture design of large-scale cloud data center based on service switch. Telecommunications Science, 2013, 29(10): 1~4

[作者简介]



赖培源,男,现就职于中国电信股份有限公司广州研究院,长期从事云计算、数据中心方面的技术跟踪研究与标准化制定工作,牵头中国电信云计算新技术引入的现网试验工作,主要研究方向为 SDN、NFV、服务器虚拟化、数据中心网络、管理平台技术、定制化服务器技术等。

马卫民,男,中国电信集团公司网络发展部高级工程师,中国电信业务网络 B 级人才,主要研究方向为电信云计算、业务网络规划建设等。

刘艺,男,现就职于中国电信股份有限公司广州研究院,主要研究方向为虚拟化、数据网络、IDC。

陈天,女,现就职于中国电信股份有限公司广州研究院,主要研究方向为互联网、云计算、IDC。