



专题:SDN & NFV

基于 SDN 的跨数据中心承载技术

刘汉江, 欧亮, 陈文华, 唐宏

(中国电信股份有限公司广州研究院, 广东 广州 510630)

摘要: 互联网产业的发展和云计算的兴起催生了跨数据中心互联的需求, 而软件定义网络技术为解决数据中心互联的网络瓶颈提供了重要途径。首先介绍了几种常用的数据中心互联技术, 分析了 SDN 的技术优势, 然后提出了一种基于 SDN 的运营商数据中心互联承载技术方案, 并阐述其实现思路。该方案可实现 DC 间的网络虚拟化, 能够提供用户业务定制、网络自动开通、任意虚拟机间的策略隔离、流量调度、用户就近访问、DC 间的容灾备份等业务功能。

关键词: 数据中心互联; 云数据中心; 网络虚拟化; SDN

中图分类号: TP393

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2016114

Bearing technology across the data center based on SDN

LIU Hanjiang, OU Liang, CHEN Wenhua, TANG Hong

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

Abstract: The development of the internet industry and the rise of cloud computing have created requirements on interconnectivity across data center, and software defined network technology provides important means to solve the network bottleneck problem of data center interconnectivity. Firstly, several commonly used interconnectivity technologies across data center were introduced, and advantages of SDN technology were analyzed. Then, the operator solution of data center interconnectivity bearing technologies based on SDN was put forward, and implementation ideas of this solution were explained. It is a solution which can realize network virtualization across data center, and provide kinds of business functions, such as service customization of user, automatic accessing of network, isolation strategy between different virtual machine, traffic scheduling, the nearest access of user, disaster recovery and backup across data center.

Key words: data center interconnectivity, cloud data center, network virtualization, SDN

1 引言

随着互联网产业的快速发展和数据中心业务的不断深化, IDC (internet data center, 互联网数据中心) 逐渐发展成为体现电信运营商核心竞争力的关键点。据全球权威咨询公司 Gartner 预测, 2016 年全球公共云服务市场规模将

达到 2 040 亿美元, 并将保持高速发展的态势, 未来几年至少 50% 的企业需要将其 IT 资源转移至云端^[1]。可见, 数据中心的云化将成为近期运营商网络建设关注的重点, 而灵活的网络架构、高效的数据处理能力、自动化的运维管理、快速的业务部署能力不仅是云计算时代对新一代数据中心的要求, 也是体现数据中心核心竞争力的基础。

目前,云计算和云存储的发展已经非常成熟,通过在数据中心的服务器上安装虚拟化软件及其管理工具即可实现计算和存储资源的池化和统一管理。但是,云服务的快速部署还需要计算、存储、网络资源的协同部署能力,网络资源需随业务申请自动按需发放,显然当前的数据中心网络难以满足以上要求。并且,在数据中心大规模部署虚拟化计算和存储以后,传统数据中心的网络问题愈加突出,逐渐成为限制其发展的瓶颈:首先,网络与业务分离,DC间的网络配置复杂,业务开通周期长;其次,4 096个VLAN限制了租户隔离资源,难以满足云计算时代多租户规模扩展的需求;再次,不同地理位置的数据中心之间三层互访,业务和资源扩展受限于单个物理DC;最后,跨DC业务需要租户租用光纤/传输资源,资源利用率低,费用高。

因此,促进跨数据中心网络互联,实现其资源的整合、池化和调度,构建一个逻辑上统一的云数据中心已经迫在眉睫。而软件定义网络(software defined networking,SDN)具有控制与转发分离、集中控制、开放可编程、网络虚拟化等技术特征,很好地契合了新一代数据中心的要求,为解决数据中心互联的网络瓶颈提供了重要途径^[2]。

本文首先介绍和分析几种常用的数据中心互联技术,然后提出一种基于SDN的运营商数据中心互联承载技术方案,并详细阐述该方案原型系统的实现架构和相应功能模块。

2 DC 互联技术

业界的DC互联技术较为丰富,大体可以分为两类:一类是基于MAC地址的二层网络互联技术,另一类是基于IP地址的三层网络互联技术。而云计算时代的跨数据中心服务器集群、虚拟机跨DC动态迁移等业务向数据中心提出了二层网络互联需求,而解决三层网络互联相对比较简单,通常可通过L3VPN实现。因此,接下来主要讨论跨数据中心二层互联技术。

2.1 光纤互联技术

基于裸光纤或DWDM(dense wavelength division multiplexing,密集型光波复用)线路的二层互联方案是通过光纤或传输资源将不同物理位置的数据中心直接互联,这种方案虽然带宽资源非常充足,网络性能比较突出,但是价格昂贵,难于扩展,主要应用于同城数据中心之间的互联。由于成本较高,对于运营商而言,该技术主要适用于高价值、对安全高度敏感的租户。

该方案主要有两种组网方式:Hub-Spoke组网和RRPP环路组网。其中,Hub-Spoke组网模型将某个数据中心的节点与多个数据中心的汇聚层节点互联,在逻辑上构成一个星型拓扑,这种组网方式基于最短路径转发,数据转发效率较高;RRPP环路组网模型将多个数据中心的汇聚层节点对等互联,所有节点之间均存在备份路径,结构强健,HA(high available)性能较优。因此,企业在选择组网方式时,应结合这两种组网模型的特点,根据应用需求进行设计^[3]。

2.2 MPLS L2VPN 技术

通过MPLS L2VPN技术,用户可以通过MPLS网络在不同站点之间透传二层数据信息。因此,可以利用该技术实现跨数据中心二层网络的延伸。目前,MPLS L2VPN技术主要有三种实现形式:VPLS(virtual private LAN service,虚拟专用局域网业务)、VLL(virtual leased line,虚拟租用线路)和E-VPN(BGP MPLS based Ethernet VPNs,以太网VPN),下面将结合DC互联的应用场景展开分析。

VPLS是目前最常见的二层互联技术,它是一种点到多点的MPLS L2VPN实现方案,多个不同地理位置的站点之间通过隧道广播学习MAC地址,对客户形成一个仿真LAN。VPLS的相关标准提出较早,成熟度高,在很多行业都广泛部署,但其组网技术复杂,存在广播域泛洪的问题,部署和运维难度高。另外,在双归属组网时主备保护负荷分担实现较为困难。

VLL是一种点到点的二层互联方式,通过IP/MPLS网络提供的VLL虚拟专线将不同物理位置的机房设备直接互联^[4]。这种互联方式路由结构清晰,但如果需要互联的数据中心节点设备较多,将存在“N的平方”问题。另外,VLL与IP网络上其他业务共享带宽资源,时延要求较高的业务可能受到不同程度的影响。

EVPN是一种用于L2域间互联的新协议,通过BGP扩展,实现远端MAC地址学习、多归属组网自动发现/冗余保护/负荷分担^[5]。该协议可减少BUM流量的泛洪,有效提高网络效率,非常适合用于数据中心互联,将成为下一代替代VPLS的L2VPN技术。但是这种技术较为复杂,配置比较复杂,各厂商支持程度低、可用性和互通性缺少验证。

2.3 overlay 技术

overlay网络方案的提出在一定程度上推动了数据中心网络的发展。该方案将数据中心网络改造为一个无阻塞的矩阵网,并通过隧道技术实现网络虚拟化和软件的集中



控制,其逻辑平面与物理平面完全分离,用户在设计 and 部署网络时无需了解底层物理网络架构。目前主流的 overlay 技术有 VxLAN(virtual extensible LAN,虚拟可扩展局域网)、NVGRE (network virtualization using generic routing encapsulation, 通用路由封装)、STT (stateless transport tunneling,无状态传输隧道)技术等,VxLAN 是当前业界的主流选择。基于 VxLAN 的 overlay 网络可以实现多个数据中心之间异构网络的二层互联,不需要考虑数据中心之间的网络类型,配置简单,是 DC 互联场景中的主推技术。

在多数数据中心互联场景中,VxLAN 通常与其他技术结合使用,如 VxLAN 与 EVPN 结合使用,EVPN 作为控制面协议,而 VxLAN 作为数据面协议。

2.4 SDN 技术

软件定义网络是一种新型的网络架构和技术,其核心理念是将传统 IP 网络的数据转发和路由控制进行分离,实现集中控制、分布转发,通过软件可编程的方式来简化和优化网络。这种控制与转发分离的网络架构使得网络不再成为制约业务快速上线和云服务效率的瓶颈,而集中控制的理念也为数据中心之间网络资源的整合、优化和灵活调度提供了重要手段。

随着 SDN 技术的发展,基于 SDN 实现数据中心互联已经成为业界认可的重要思路,集中的 SDN 控制平台也逐渐成为实现数据中心之间流量调度和网络资源自动化管理的关键部件。

在 DC 互联场景中,SDN 技术应用的主流思路有两种:一种是“SDN+ VxLAN”的 overlay 网络解决方案,另一种是“SDN+openFlow”的 underlay 网络解决方案。前者以 VxLAN 作为数据面协议,标识租户,并透明传输租户信息,SDN 控制器集中管理和控制各个数据中心的 VxLAN 隧道端点设备,并根据业务需求向相应设备下发指令,对进出数据中心或服务器的报文进行 VxLAN 封装转换;后者需要数据中心的网络设备支持 OpenFlow,SDN 控制器根据业务策略集中生成流表,并通过标准化的 OpenFlow 协议下发给数据中心的 OpenFlow 设备,OpenFlow 设备只需根据流表对匹配到的数据分组执行相应的转发指令即可。

在这两种思路中,基于 VxLAN 的 SDN 方案相对较为成熟,厂商支持程度较高,可作为近期实现数据中心互联的过渡技术。但由于基于 overlay 的 DCI (data center interconnection,DC 互联)虚拟网络可能跨广域网,而

VxLAN 对网络状况并无感知,因此难以规避拥塞,无法提供有效的流量调度和 QoS 保障。而基于 OpenFlow 的 SDN 方案可以从中央控制节点获取网络配置信息,感知虚拟网络内跨 DC 的流量访问,根据业务类型对某些时效性要求高的流量进行保障传送,是未来数据中心网络演进的重要方向。

3 运营商 DC 互联承载技术方案与实现

目前,基于 SDN 技术实现跨数据中心网络的互联已经成为目前 DC 互联的主流思路。在这种背景下,本文提出一种 SDN+MPLS 的解决方案,实现运营商跨数据中心网络的多租户互联。

3.1 目标与需求

为应对云计算时代业务弹性扩展、就近接入、跨地域容灾备份、DC 间数据同步、资源动态调度、虚拟机迁移等业务需求,未来的 DC 互联网络存在 3 个发展目标。

- 资源虚拟化:实现 DC 间的网络虚拟化,对用户呈现逻辑上统一的云数据中心,实现数据中心之间资源的整合、池化和调度,如图 1 所示。

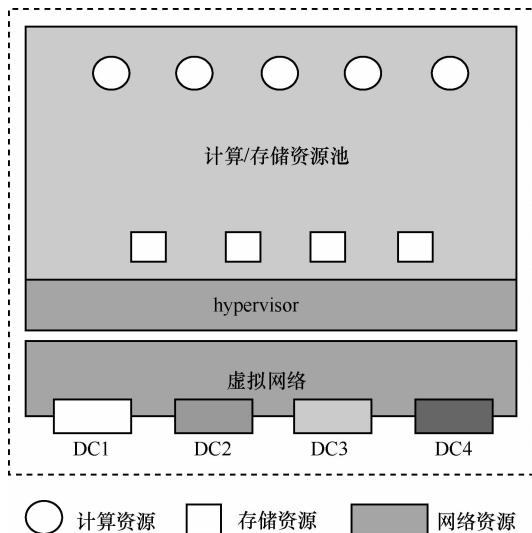


图 1 虚拟化数据中心示意

- 业务自动化:实现网络自动配置,业务灵活部署。
- 流量可调度:提供多条 DC 间的业务路径,用户可根据需要灵活选择和调整业务流的路径。

为满足云服务的网络需求,运营商 DC 互联承载技术方案的总体功能简述如下。

- 实现灵活的二层虚拟网络和跨 DC 虚拟化网络的延伸。

- 面向用户提供定制界面,完成虚拟网络配置服务,实现网络的自动开通。
- 实现任意虚拟机的策略隔离。
- 提供多种业务路径,用户可根据实际需要动态调整业务流的路径。
- 实现用户跨 DC 虚拟网络的分布部署,如 Web 靠近用户前置部署,使用户可就近访问,缩短用户访问的不稳定性。

3.2 技术方案

根据上述功能要求,本节提出一种“SDN+MPLS”的 DC 互联方案,其组网架构如图 2 所示,整个系统可分为 3 层:承载层、控制层和协同管理层,下面将对该架构进行说明。

承载层由数据中心和骨干网的网络设备组成,包括 DC 内的 OpenFlow 交换机(接入和汇聚交换机均采用 OpenFlow 交换机)和服务器的骨干网的路由器等。数据中心的服务器可以安装商用或开源虚拟化软件,这里假设采用 VMware 的虚拟化方案,其 hypervisor 为 ESXI,虚拟机之间通过虚拟交换机 VDS 相连。

控制层由 DC 控制器和 vCenter 组成。其中,DC 控制器负责根据协同器下发的策略进行流表映射,并下发流表表项,实现对 OpenFlow 交换机的控制;对于出入接入交换机的数据分组弹出/压入 VLAN 标签,实现物理主机内不同租户的隔离;并压入/弹出所选 LSP 的 MPLS 标签栈,实现路径选择,内层标签表示目的 DC 内的 LSP 路径,中层

标签表示 DC 间骨干网的 LSP 路径,外层标签表示源 DC 内的 LSP 路径。在整个 DC 互连网络,只根据 MAC 地址全局判定租户虚拟网络。

协同管理层由协同器和 DCI 云管理平台组成。DCI 云管理平台负责 DCI 的网络管理和控制功能,并面向用户提供虚拟网络配置服务。协同器接收云管理平台的指令,负责协调和管理各 DC 的控制器,并向各 DC 内的控制器下发业务策略。

3.3 基于 MPLS 的 DC 互联承载方案实现

考虑采用图 3 所示功能架构,各层次间的相关处理流程简述如下。

- 用户通过 DCI 云管理平台定制其虚拟网络,包括创建业务主机、选择路径策略等。
- DCI 云管理平台通过 vCenter 将业务主机实例化,vCenter 将业务主机的部署信息反馈给 DCI 云管理平台。
- DCI 云管理平台将业务主机的部署信息与物理网络拓扑的相关信息存储至用户信息数据库,并透传给协同器。
- 协同器根据用户定义的网络关系和路径策略进行路径计算,基于当前的网络资源状况选择 LSP 路径,并将相应的策略模板下发给对应的 DC 控制器。
- DC 控制器将路径策略映射为流表,下发给对应的 OpenFlow 交换机。

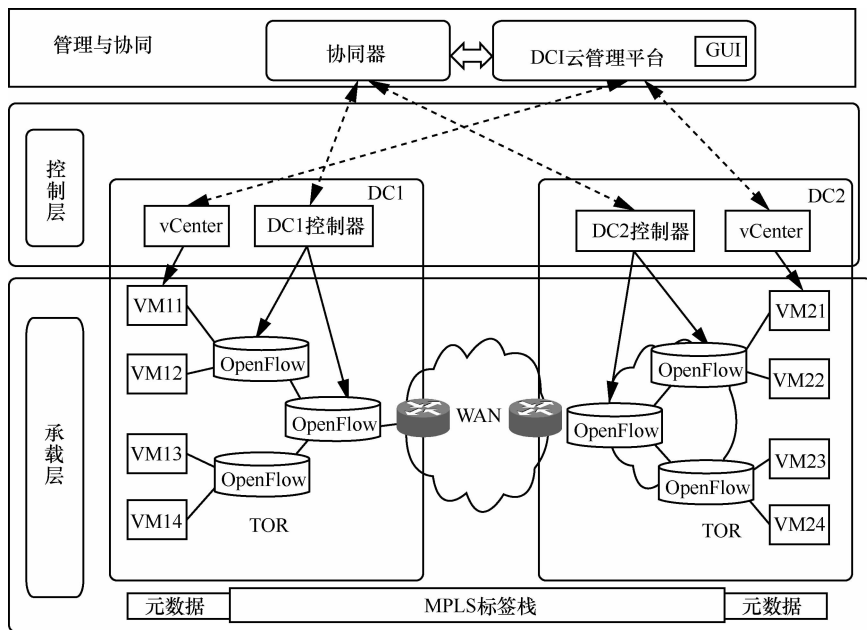


图 2 运营商 DC 互联承载技术方案组网架构

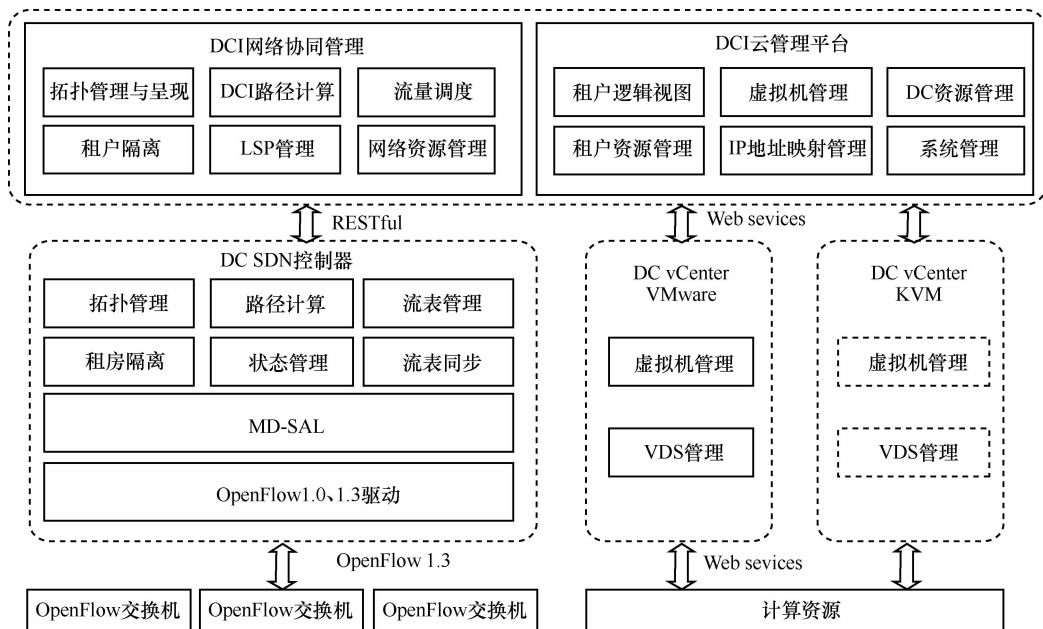


图3 系统功能架构

- OpenFlow 交换机接收并匹配到指定业务报文,则按照指定策略,执行相应的转发处理同时,交换机将端口流量的统计数据上报给 DC 控制器,并由后者通过 RESTful 接口上报给协同器。
- vCenter 将 DC 计算资源的运行状况上报给 DCI 云管理平台。

下面介绍主要功能模块。

3.3.1 DCI 云管理平台功能

DCI 云管理平台上具备面向运营商、企业、租户的 portal,提供业务灵活定制的图形化界面,并负责 DCI 的资源管理,具体功能如下。

- 租户逻辑视图:保存和维护租户所定义的逻辑网络视图。
- 租户资源管理:管理租户需要使用的资源,包括租户可用 DC、计算资源、VLAN、IP 地址、MAC 地址、账号和密码等。
- 虚拟机管理:创建、删除虚拟机,并管理虚拟机模板(业务逻辑)。
- IP 地址映射管理:主要管理公网 IP 地址和租户私网 IP 地址之间的映射关系。
- DC 资源管理:与各 DC vCenter 联动获取 DC 的计算资源情况并进行管理。
- 系统管理:包括用户、管理员的权限管理,操作日志管理等。

3.3.2 协同器功能

协同器又称为 DCI 网络协同管理模块,负责总体的业务编排及多厂商 DC 控制器的协同,具体功能如下。

- 拓扑管理与呈现:网络物理拓扑的管理与维护,并将拓扑透传给 DCI 云管理平台,在图形化界面上展示。
- 租户隔离:实现物理主机内的租户隔离。本方案全局使用 MAC 地址池实现租户隔离,在物理主机内使用 VLAN 隔离租户。VLAN 在同一个物理主机内有效,不同物理主机 VLAN 可复用,成功解决 DC 隔离资源受限的问题。同租户不同物理主机间的 VLAN 转换通过租户标识、VM MAC 地址、VLAN 和物理主机 ID 等信息表实现。
- DCI 路径计算:根据用户在云管理平台上指定的路径策略进行 LSP 路径计算。
- LSP 管理:管理和维护 LSP,包括 MPLS 标签、跳点、LSP 策略模板等。
- 流量调度:负责实现 DC 间的交换机/租户/VM 之间的流量调度。
- 网络资源管理:管理、分配、回收租户使用的网络资源,包括交换机的管理 IP 地址、路由器的 MAC 地址、端口和连接关系等。

3.3.3 控制器功能

DC 控制器负责网络建模和网络实例化,将协同器下

发的策略进行流表映射,并下发给 OpenFlow 交换机,将底层资源状况反馈给协同层,主要基于 OpenDaylight 开源控制器实现,具体功能如下。

- 拓扑管理:负责 DC 内物理拓扑的管理与维护,并将数据中心的物理拓扑发送给协同器。
- 路径计算:根据协同器的路径计算模块下发的路径策略模板生成流表。
- 流表管理:负责流表的下发、删除、修改、查询,与 OpenFlow 交换机的流表同步。
- 状态管理:管理 OpenFlow 交换机之间、OpenFlow 交换机与 DC 控制器之间的连接状态。
- 租户隔离:负责维护本数据中心内租户标识,如虚拟机 MAC 地址、VLAN 和物理主机 ID 等信息,实现不同租户之间的隔离。
- MD-SAL:OpenDaylight 提供的基于模型驱动的服务抽象层,为上层的功能模块提供一致性服务。
- OpenFlow 1.3 驱动与接口:支持与 OpenFlow 1.3 交换机通信。

3.3.4 业务流程

通过上述功能开发,该方案可实现 DC 间的网络虚拟化,能够提供用户业务定制、网络自动开通、任意虚拟机策略隔离、DC 间流量调度、用户就近访问、DC 间的容灾备份等业务功能。由于系统功能较多,这里不一一赘述,仅以两个重要功能为例来说明其实现原理。

(1) 网络自动开通

首先,租户在 DCI 云管理平台上的 GUI 上创建业务主机并通过 vCenter 将其实例化,vCenter 将业务主机的部署信息反馈给 DCI 云管理平台(包括部署 DC 和所属物理主机 ID)。DCI 云管理平台将业务主机部署信息与物理网络信息进行整理,存储至用户信息数据库,如图 4 所示。

VM ID	VM IP 地址	VM MAC 地址	主机 ID	所属 DC	交换机 ID	交换机端口号
-------	----------	-----------	-------	-------	--------	--------

图 4 用户信息数据库

用户完成业务主机的创建工作之后,需要为其定义网络关系,包括子网间的互通关系,主机之间互访关系等,并选择 DC 内、DC 间虚拟机互通的路径选择策略。DCI 云管理平台将相关网络策略和用户信息数据库一起透传给协同

器。协同器根据租户定制网络和当前的网络资源状况进行路径计算,选择 LSP,并向对应的 DC 控制器下发策略。DC 控制器根据路径策略集中生成流表表项,下发给对应的 OpenFlow 交换机。OpenFlow 交换机根据流表转发数据分组。

至此,便实现了租户网络的自动开通。云管理平台、协同器、DC 控制器、OpenFlow 交换机之间的关系如图 5 所示。

(2) DC 间流量调度

网络管理员预先静态配置多条 DC 间的 LSP 路径,包括默认路径、高带宽 VIP 通道、其他带宽路径等。

租户可以根据业务需求调整 DC 间的路径,实现业务流的动态路径优化。

租户在 GUI 上重新选择 DC 间的业务路径后,DCI 云管理平台向协同器下发变更 DC 间 LSP 的请求,协同器收到请求后分别向源 DC 和目的 DC 下发路径变更策略。控制器删除原有 Openflow 交换机中相应的流表组,并根据新的 DC 间路径重新生成和下发新的流表组。

4 结束语

促进跨数据中心的网络互联,整合全网的 DC 资源是云计算时代数据中心的发展趋势。文中对光纤互联、L2 VPN、overlay 和 SDN 等多种主流 DC 互联技术进行了分析。其中,基于 SDN 技术实现跨数据中心网络的互联已经成为业界认可的重要思路,集中的 SDN 控制平台也成为了实现业务全网管控和网络资源自动化管理的关键部件。

在这种背景下,本文阐述了一种运营商 DC 互联承载技术方案及其实现方法。该方案可实现 DC 间的网络虚拟化,能够提供用户业务定制、网络自动开通、任意虚拟机策略隔离、流量调度、用户就近访问、DC 间的容灾备份等业务功能。

参考文献:

- [1] 199IT 中文互联网数据资讯中心 [EB/OL]. <http://www.199it.com/archives/435019.html>. 2016.
- [2] 199IT Chinese internet data information center [EB/OL]. <http://www.199it.com/archives/435019.html>. 2016.
- [2] 唐宏,刘汉江,陈前锋,等. OpenDaylight 应用指南[M]. 北京:人民邮电出版社,2016.
- TANG H, LIU H J, CHEN Q F, et al. Application

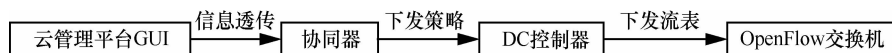


图 5 网络自动开通业务流程示意



guidelines of OpenDaylight[M]. Beijing: Posts & Telecom Press, 2016.

- [3] 马骥. 基于 SDN 的云数据中心互联[J]. 电信工程技术与标准化, 2015(11): 54-59.

MA J. Cloud data center interconnect based on SDN[J]. Telecom Engineering Technics and Standardization, 2015(11): 54-59.

- [4] 姚金玲. 企业数据中心二层网络互联技术研究 [J]. 网络安全技术与应用, 2015(5): 85-86.

YAO J L. L2 network interconnection technology research on enterprise data center [J]. Network Security Technology & Application, 2015(5):85-86.

- [5] BGP MPLS Based Ethernet VPN: IETF. Draft-ietf-l2vpn-evpn-03[S]. [S.l.: s.n.], 2013.

[作者简介]



刘汉江(1988-),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IP承载网络技术、SDN/NFV技术。



欧亮(1968-),男,博士,中国电信股份有限公司广州研究院高级工程师,主要研究方向为电信网络规划设计、互联网新技术研究与应用。



陈文华(1967-),男,中国电信股份有限公司广州研究院高级工程师,主要研究方向为数据通信网络。



唐宏(1974-),男,中国电信股份有限公司广州研究院数据通信研究所所长,主要研究方向为IP承载网、下一代互联网和网络新技术方面的研发与管理。