



基于业务交换机的大规模云数据中心通用网络架构设计

樊勇兵, 丁圣勇, 陈楠

(中国电信股份有限公司广东研究院 广州 510630)

摘要:业界首次提出业务交换机的概念,基于业务交换机的大规模云数据中心通用网络架构,可以为数据中心网络设计提供通用方法论,从而简化设计并使逻辑清晰,帮助运营商提高运营灵活性、降低成本与风险,为设备商提出一种可能的设备形态以支持这种架构。

关键词:业务交换机;云计算;数据中心;云数据中心

doi: 10.3969/j.issn.1000-0801.2013.10.001

A General Network Architecture Design of Large-Scale Cloud Data Center Based on Service Switch

Fan Yongbing, Ding Shengyong, Chen Nan

(Guangdong Research Institute of China Telecom Co., Ltd., Guangzhou 510630, China)

Abstract: The concept of service switch was firstly proposed around the industry. The general network architecture of large-scale cloud data center based on service switch can provide a general methodology to simplify the design and to clarify the logic of data center network, which will help carrier improve flexibility, lower cost and risk of operation, and a possible device form for vendor to support the architecture was suggested.

Key words: service switch, cloud computing, data center, cloud data center

1 大规模云数据中心简介

1.1 概述

数据中心(data center, DC)一般粗略地分为互联网数据中心(internet data center, IDC)和企业数据中心(enterprise data center, EDC)两大类,前者指对个人或集团客户公开提供托管、租赁服务的数据中心,后者指企业自用数据中心(暂将Google等互联网公司用于内部后台计算的数据中心也归为此类)。本文的研究对象是前者的总体网络组织。

数据中心的规模一般用所能容纳的服务器数量衡量。业界盛传Google数据中心有上百万台服务器,分布在数个大型数据中心。以此推之,每个数据中心可能有数十万台

服务器。每个数据中心一般由多个机房或机楼组成(暂不考虑集装箱数据中心),所以单个机房应有数万台服务器;截至2012年底,国内已投产规模最大的数据中心,其规模也在5~8万台服务器。粗略地将单一管理域内每机房具有1万台以上服务器、整个园区具有5万台以上服务器的数据中心称为“大规模”数据中心。

上面所指的服务器均为物理服务器。如不特别指出,本文所称服务器不区分物理服务器和虚拟服务器(又称为虚拟机)。

所谓云数据中心,指一个数据中心由云计算技术构建并为客户提供云计算服务。

目前,整个数据中心产业正处于从传统模式向云计算模式过渡的早期阶段,传统数据中心有很大的存量资产,



云计算涉及领域众多,流派各异,技术和标准都不成熟,运营商和设备商都面临选择的难题。为大规模云数据中心提出一个通用参考网络架构,以提高运营灵活性、降低成本与风险,并为设备商提出一种可能的设备形态以支持这种架构,是本文希望达成的目标。

1.2 云数据中心的业务特性

云业务就是将弹性、池化的 IT 资源作为服务通过互联网提供给用户,云数据中心是这种服务的承载体,其业务特性主要表现在以下几个方面。

(1) 商业模式

云数据中心的业务模式从以托管和固定计费为主转向以租赁和按使用计费为主,业务开通从以线下为主转向以线上为主,用户类型从以企业为主转向企业、个人并重。

(2) 资源类型

由于资源的云化和可迁移,云数据中心的资源类型从以专用物理资源为主转向以可弹性复用的虚拟资源、业务资源为主,业务接入则从以固定为主转向固移结合。

(3) 业务规模

资源云化使大规模业务提供成为可能。以一个 5 万台物理服务器的数据中心为例,假设每台服务器虚拟为 15 个虚拟机(VM),给每个 VM 分配 2 个 MAC 地址和 1 Mbit/s 带宽,则所有服务器的 MAC 地址数量为 150 万个,出口带宽为 750 Gbit/s。如果考虑到下行连接、链路冗余和一定的网络裕量、横向流量,则核心层网络设备的实际端口容量大致为 2.5 Tbit/s。这只是一般规模下非常保守的估计,但已经比传统数据中心的交换机容量提高了一个量级。

(4) 业务流量

流量大小:其量级在上文已有论述。在这样的量级上,流量复用效率可以明显改善,但流量的协调、伸缩和不可预见性则需要更好地管理。

流量方向和交互:传统数据中心的流量方向主要是南北向(又称为纵向)的,即数据中心内部和外部之间交互流量;云数据中心则有大量东西向(又称为横向)流量,即数据中心内部服务器之间交互流量。东西向流量主要来源于以下 3 个方面:

- 新型计算模式,如 Hadoop 计算所产生的流量;
- 灾备、存储迁移、虚拟机迁移流量;
- 在公有云中,由于大量用户将自己的 DC 建立在云服务提供商(cloud service provider,CSP)提供的基础设施上,则用户的业务交互对基础设施而言是可

感知而不可预见的,这不同于传统数据中心。

流量组成:除了传统数据中心已有的管理流量、控制流量、业务流量、备份流量等,云数据中心还有自己特有的迁移流量、虚拟机镜像流量。另外,云数据中心的业务流量所承载的用户数远大于传统数据中心,控制流量远比传统数据中心的丰富、复杂。

1.3 对云数据中心网络的要求

综上所述,云数据中心与传统数据中心的业务特性差异巨大,这种差异在数据中心网络层面主要表现在以下 4 个方面。

(1) 网络边界

在云数据中心,由于服务器虚拟化技术的采用,网络已经延伸到服务器内部,因此,物理网络和虚拟机、虚拟交换机(vSwitch)之间需要互操作、互感知。有两种可能的解决方案:一种是通过网络上的管理平台,一种是在服务器和交换机之间建立协议关系,如 IEEE 802.1Qbg、IEEE 802.1BR。

(2) 网络层次

云数据中心要接入海量的互联网用户以及大量的服务器,前者要求数据中心网络有中等或以上的路由能力(视网络结构而定),后者要求数据中心网络有大二层能力。所谓“大二层”,综合体现在服务器或 VM 数量、MAC 地址数量、网络设备容量、网络流量和用户数量等方面,并且有弹性伸缩、广域互联、细颗粒度负载分担等需求。大二层解决方案主要有两种:一种是广域方案,可以称为数据中心互联(data center interconnect,DCI),如 VPLS、PBB、PBB-EVPN 等;另一种是园区级或城域级方案,如叠加(overlay)型或隧道(tunnel)型、交换型(如 trill)方案。上述方案中有些内置了冗余和负载分担等功能,还有些属于设备和链路级的冗余和负载分担方案。

(3) 流量模型

在云数据中心,南北向流量和东西向流量兼备,这些流量将对网络上下行收敛比、设备的架构和端口缓存设计产生重大影响;另外有一些网络服务,如防 DDos 服务,需要 VPN 功能回送清洁流量到目的设备;由于服务器的虚拟化,有些传统网络服务需要下沉到二层甚至物理服务器内部实现。

(4) 流控

云数据中心规模巨大,用户众多,流量复杂,需要对不同用户、不同属性的流量进行区域、路径、QoS 等控制。

1.4 传统数据中心组网的局限性

传统数据中心的组网特点基本可以概括为“一房一网,分层汇聚”。一房一网,是指数据中心的网络以机房为单位,一个机房就是一个独立的网络,每个机房都连接到骨干网或城域网,甚至作为城域网的一部分,各机房之间没有关系;分层汇聚,是指数据中心网络基本按接入、汇聚、核心、出口 4 层进行组织,每个层次的核心作用雷同,基本是汇聚下层、连通上层以及对流量进行南北向转发,外加一些简单的冗余、防环机制。

显然,这样的架构已经不足以支持云业务。

2 业务交换机概述

云计算对数据中心网络的要求赋予数据中心网络设备丰富的业务特性,略举如下:

- 北向三层路由和两层交换的分界;
- 南向接入时对虚拟机、虚拟交换机甚至存储的感知与交互;
- 横向大二层能力和作为二/三层 DCI 的客户边缘(customer edge, CE)设备能力;
- 必要的 VPN 能力;
- 各层次网络服务(如安全、负载均衡等)的提供(可能作为交换机外挂的物理设备,也可能作为交换机的硬件组成);
- 均衡的南北向流量和东西向流量转发能力;
- 复杂的流控能力;
- 适当规模的路由能力、交换容量、MAC 地址支持能力等。

如何在通用方法论的指导下组织上述复杂的业务和技术特性,使之形成有机整体,并能适应多流派(如大二层问题的叠加方案和交换方案)、多制式(如服务器虚拟化的多种技术)、多业务(如传统业务和云业务)的共存,规模上可扩展并可分步实施,技术上可平滑演进,防止厂商锁定,简化数据中心网络设计,是“业务交换机(service switch, SS)”要解决的问题。

所谓“业务交换机”,是一种数据中心网络设备,集成了数据中心的核​​心业务能力(物理上可以适当分离),北向提供互联网用户接入,南向提供数据中心资源接入,横向提供业务交互和扩展能力,交换机起到边界确定、连接分类、规模扩展、业务控制、流量转发的作用。从数据中心的整体看,所有业务交换机共同构筑了一个边界层,是整个

网络的灵魂。

3 以业务交换机为基础的云数据中心通用网络架构

3.1 设计思路 and 要点

以业务交换机为基础的云数据中心通用网络架构,其基本设计思路是:功能集成,逻辑清晰,分层分模块组织。要点如下:

- 将整个数据中心网络划分为骨干层和业务层;
- 骨干层主要为业务层提供数据中心出口服务、长途 DCI 服务和骨干层网络服务,逻辑上对应的设备形态分别是出口路由器、DCI PE(provider edge)、骨干层网络服务设备;
- 业务层的主要功能是通过业务交换机实现北向用户接入、南向业务承载、横向业务交互和扩展、业务层的网络服务,在业务层内,以业务交换机为核心组织业务块(service block),若干类似属性的业务块经区域交换机汇聚为业务区域(service domain),若干业务区域(可能是不同属性的)共存于一个业务空间(service space),所有业务空间构成数据中心网络的业务层,并与骨干层互联;
- 区域交换机的作用主要是对业务交换机进行汇聚,如果规模足够小或业务交换机足够强大,可以取消区域交换机,以使网络更加扁平。

3.2 设计示例

通用网络架构设计示例如图 1 所示,整个云数据中心的骨干层出口路由器分别与公众互联网、高等级互联网(或长途光网、DCI 专网)连接,DCI PE 是 DCI 网的边缘设备,trill 交换机可实现 trill 方式的大二层互联,亦可作为汇聚交换机或互联交换机连接各业务交换机或有需要的相关网络设备;业务层可划分为若干业务空间(物理上可能对应若干机房、机楼),每个业务空间划分为自主客户业务区、传统托管业务区、云业务区,其中云业务区又可划分为个人业务块、企业业务块;园区范围内的光网(或裸纤、交换)用于园区内部的物理互联,可承载业务交换机之间的大二层连接和到 DCI PE 的连接、区域交换机到出口路由器的连接、存储灾备等。

典型业务区块的业务特征和某些技术考虑如下。

(1) 自主客户业务区

数据中心运营商只提供空间、水电、散热、骨干网络接

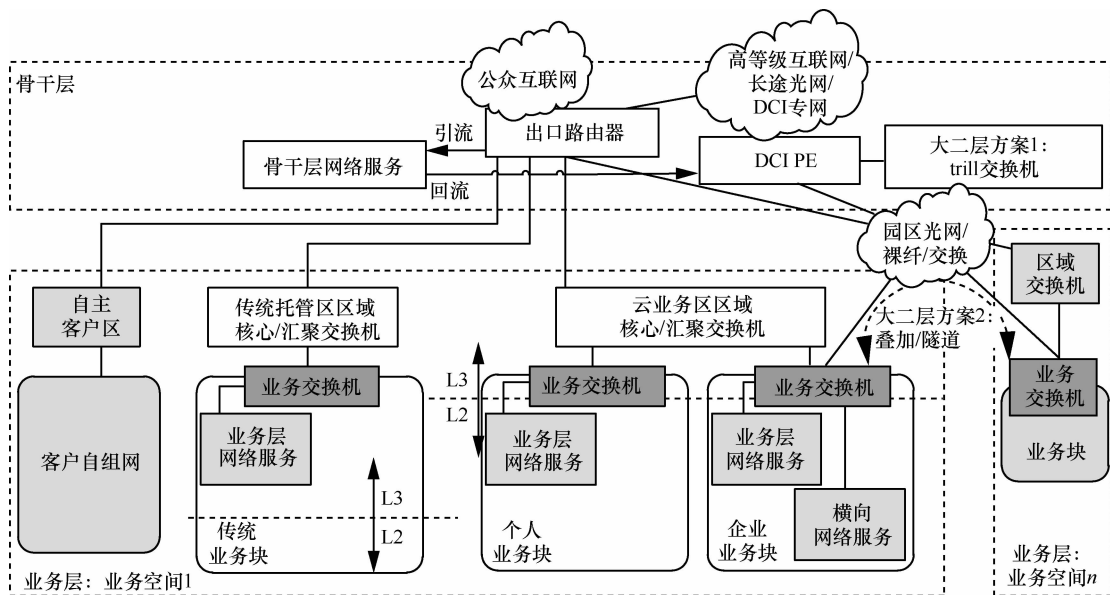


图1 通用网络架构设计示例

入、(可选的)传输;用户自己组网(典型用户如互联网公司、企业租赁用户)。

(2)传统托管业务区

在自主业务区的基础上,运营商还需提供机房内的网络接入和(可选的)网络服务(如安全、负载均衡);用户提供服务器、存储,有时也会自带接入交换机(即 top of rack, ToR)、防火墙等网络设备;以南北向流量为主;DCI需求以三层VPN为主,可将业务交换机作为CE接入DCI PE,亦可专设区域CE或业务块CE接入PE;该区域的业务交换机可使用传统交换机,直接接入IT设施或者通过ToR接入。

(3)云业务区

运营商提供所有基础设施(甚至可能包括应用)和各层次网络服务(包括可能的横向网络服务),用户只购买服务(面向个人用户提供标准云业务,面向企业用户提供虚拟数据中心(virtual DC, vDC));南北向和东西向流量共存;业务交换机之间通过trill或叠加等方式实现园区大二层,并通过DCI PE接入广域大二层或三层VPN;该区域的业务交换机是本文所定义的典型业务交换机,可提供超高密度10 Gbit/s端口,通常直接接入服务器和存储。

4 结束语

云计算给数据中心带来了很复杂的挑战。通过本文定义的业务交换机,可以为数据中心网络设计提

供通用方法论,简化设计,并使逻辑清晰。但也看到,逻辑清晰并不意味着逻辑简单——相反,云数据中心(包括本文并未涉及的互联网公司的数据中心)涉及面太广,现有业务和技术都太复杂。也许下一步追求的是一个逻辑清晰简单的解决方案。随着硬件技术的进一步发展,软件定义网络(software defined networking, SDN)可能会带来惊喜。

[作者简介]



樊勇兵,男,博士,中国电信股份有限公司广东研究院高级工程师,全球云网 Open Cirrus 中国电信节点负责人,中国电信重点实验室——数据多媒体实验室负责人,中国数据中心产业发展联盟云计算服务专家委员会第一批专家委员,中国通信学会高级会员,长期从事互联网和云计算研究工作,目前负责中国电信云计算、IDC等方面的工作。

丁圣勇,男,中国电信股份有限公司广东研究院工程师,长期从事互联网技术的应用研发工作,主要研究方向为网络大数据挖掘分析。

陈楠,男,博士,现就职于中国电信股份有限公司广东研究院,ITU-T 云计算基础设施即服务功能需求及架构(Y.ccIaaS)编辑人,目前主要从事云计算及数据中心方向的研究工作。

(收稿日期:2013-09-01)