

叶脊架构在数据中心的应用

Application of Spine-Leaf Architecture in Data Center

包琅允(湖南省邮电规划设计院有限公司,湖南长沙410126)

Bao Langyun(Hu'nan Planning & Designing Institute of Post & Telecommunication Co.,Ltd.,Changsha 410126,China)

摘要:

目前,互联网数据中心数据量主要集中在PB/EB级别,在采取虚拟化技术之后,东西流量大幅增加,给传统三层架构带来很大挑战。叶脊架构将网络从三层扁平化为二层,更适于数据中心内部流量的高效运转,帮助企业充分发挥网络的优势,寻求数据中心效用的最大化,提升资源利用率。叶脊架构对整个高端光模块的应用数量是传统架构的数十倍,将强势拉动高端光模块和光器件需求,进一步提升网络容量并降低每比特成本,可有效解决业务流量以及网络带宽持续增长的压力。

Abstract:

At present, the data volume of Internet data center is mainly concentrated in PB/EB level. After the adoption of virtualization technologies, the east-west traffic increases significantly, which brings great challenges to the traditional three-tier architecture. The spine-leaf architecture flattens the network from three layers to two layers, which is more suitable for the efficient operation of internal traffic in the data center. It helps enterprises give full play to the advantages of the network, seek the maximum utility of the data center, and improve the resource utilization. The number of entire high-end optical module applications in spine-leaf architecture is tens of times as great as the traditional architecture, which will strongly drive the demand for high-end optical modules and optical devices, further increase network capacity and reduce cost per bit, effectively solve the growing pressure in service traffic and network bandwidth.

Keywords:

IDC; Spine-leaf architecture; Flat network

引用格式:包琅允. 叶脊架构在数据中心的应用[J]. 邮电设计技术,2021(2):74-77.

0 前言

云计算是信息技术发展的主流,政府将它视作信息化和工业化结合的支撑点;运营商把它作为超越管道的转型引擎;企业和行业把它看成是构建智能和敏捷的现代化企业征程中一场新工业革命。

从产品与应用视角来看,需要建设大量的数据中心(IDC),通过数据中心对资源集中部署和共享,实现提升效率、降低成本的发展目的。通过数据中心开放性生态平台提供丰富的ICT业务,以此促进各个行业

关键词:

数据中心;叶脊架构;扁平化网络

doi:10.12045/j.issn.1007-3043.2021.02.014

文章编号:1007-3043(2021)02-0074-04

中图分类号:TN914

文献标识码:A

开放科学(资源服务)标识码(OSID):



的数字化、信息化和智能化,为“互联网+”、新基建相关行业的发展提供有力支撑,同时激发出更多新增需求。

1 传统数据中心胖树架构

目前数据中心大多采用胖树(Fat-Tree)架构,使用大量的高、中、低性能交换机,构建出大规模的无阻塞网络。胖树架构中,网络带宽不收敛,保障在任意的通信模式下,通信带宽达到网卡带宽。采用胖树架构搭建数据中心之后,数据中心实质是一种传统的三层硬件架构(见图1)。

接入层:用于连接所有的计算节点。通常以机柜

收稿日期:2020-12-16

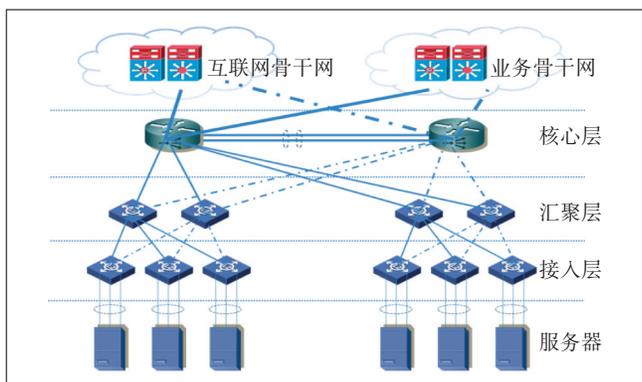


图1 传统三层架构示意图

交换机(TOR)的形式存在。

汇聚层:用于接入层的互联,并作为该汇聚区域二层和三层网络的分界点。同时提供防火墙、SSL offload、入侵检测、网络分析等其他服务。

核心层:为进出数据中心的包提供高速的转发,用于汇聚层的互联,同时实现整个数据中心与外部网络的三层通信。

表1示出的是传统架构交换机结构。

表1 传统架构交换机结构

类目	层级	性能相对要求	功能
核心层交换机	L3	高	提供弹性的三层路由
汇聚层交换机	L2/L3	中	二层和三层之间的策略承载
接入层交换机	L2	低	终端业务区分识别和接入

这种传统的数据中心网络架构,设备独立,单台设备部署单项功能并单独管理,服务器、交换机、路由器、网络、安全等设备可由不同厂商提供。数据中心企业在部署内部机架连接,特别是短距离连接时,大多数都是使用铜缆布线连接。目前6类、6A类万兆铜缆系统仍然是数据中心铜缆布线系统的首要选择。在这种架构中,铜缆布线使用率达到87%,而光缆使用率只占13%。

但这种传统的三层架构有如下明显的缺陷。

a) 硬件资源利用率极低,硬件投资成本大,管理维护难。

b) 采用生成树协议(STP),实际承载流量的只有一条,其他上行链路只用于备份,是被阻塞的,带宽资源浪费明显。

c) 故障域大、排障难,STP协议由于其本身的算法,在网络拓扑发生变更时需要重新收敛,容易发生故障,影响整个VLAN网络。

d) 随着数据中心机房中网络设备集约化程度的不断提升,铜缆布线带宽及高密度未来模块化升级难度大,在数据中心布线中光纤逐步成为主导地位。

e) 数据中心当前大量使用的铜缆介质,不符合绿色节能可持续发展要求。

f) 最为重要一点,随着数据中心采用虚拟化技术和微服务架构后,平级设备之间的数据量大幅增加,传统三层架构核心交换机和汇聚交换机的端口和性能将遭遇瓶颈。

2 新型叶脊架构在数据中心应用

从数据中心发展趋势来看,企业上“云”、数字化的趋势不可逆转,云计算、大数据、人工智能等信息通信技术(ICT)行业对数据中心的需求量越发庞大,数据中心带宽需求正在以每年25%到35%的速度飞速增长,2020年,带宽容量会将翻倍,达到50G至100G。

相比于传统的数据中心数据流,云计算数据中心数据流主要在数据中心服务器之间内部流动,是一种东西向流动。这种东西走向的大量数据的交换要求数据中心采用扁平化的网络架构,这种新型“胖树”架构能够实现无阻塞、低延迟、快速交换,也促使网络布线结构做相应的改变。

正是基于上述背景,提出一种改进型胖树架构即叶脊网络架构,和传统数据中心胖树架构一样,这种新型胖树架构同属于CLOS网络模型,但不同的是,叶脊架构采用大二层网络架构,扁平化为骨干和分支两层(见图2)。

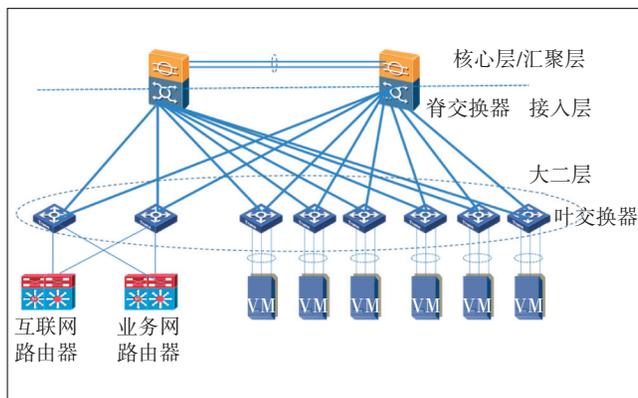


图2 叶脊架构示意图

叶脊网络大二层架构可以满足云计算数据中心安全性、易用性、灵活性、可用性,二层和三层的分界在脊交换机,主要利用大二层中的叶交换机搭建整个数据

中心,部署独立的二层广播域。叶脊架构中,所有横向的服务器在网络位置上是平行的,网络效率高,传输通道完全独立,非常适合高性能计算集群和高频流量通信环境。

叶交换机,相当于传统三层架构中的接入交换机,作为机柜交换机直接连接服务器、存储设备、防火墙、负载均衡器、边界路由器等终端设备。叶交换机之间的服务器的通信,需要经由脊交换机进行转发。

脊交换机,相当于传统三层架构中的核心交换机,叶和脊交换机之间通过等价多路径路由协议(ECMP)动态选择多条路径与叶交换机层的各个交换机连接。

表2示出的是叶脊架构交换机结构。

表2 叶脊架构交换机结构

类目	层级	性能相对要求	功能
脊交换机	L3/L2	较高	网络的主干,实现路由
叶交换机	L2	低	连接终端设备

数据中心作为新基建一项重头,有必要考虑部署40G及以上高带宽网络。随着标准的成熟和技术的提升,200/400G高速率光模块的商用,为叶脊网络在数据中心的部署提供了可行性。叶脊架构相比传统三层硬件架构,更能符合数据中心发展趋势,其优势体现在:

a) 均衡分担负载,每个叶交换机的上行链路以负载均衡方式工作,充分利用了带宽。

b) 预估延迟值,叶交换机之间的连通路路径的条数可确定,均只需经过一个脊交换机,东西向网络延时可预测。

c) 扩展性好,大多数支持25G/100G的交换机都能向后兼容10G/40G,为交换机的端口提供更高的灵活性,网络扩容更加灵活。

d) 降低交换机的性能要求,南北向流量,可以从叶节点出去,也可从脊节点出去。东西向流量,分布在多条路径上。这样一来,不需要昂贵的高性能高带宽交换机。

e) 高安全性和可用性,叶脊架构中,一台设备故障时,不需重新收敛,流量继续在其他正常路径上通过,网络连通性不受影响,带宽也只减少一条路径的带宽,性能影响微乎其微。

在叶脊架构下进行设计时,中小型数据中心超过90%的光纤链路长度小于100m,采用多模光纤为主的链路部署方案,光模块部署以40/100G为主。大型

数据中心超过70%的光纤链路长度小于100m,超过85%的长度小于150m,也是考虑采用多模光纤为主的链路部署方案,可以前瞻性地考虑部署200/400G光模块。

以一个中小型数据中心为例,叶脊网络的支撑能力设计如表3所示。

表3 叶脊网络支撑能力设计

项目	类目	参数	端口数量/个	光模块
IDC	脊交换机	上联端口	$S_{上}$	100G
		下联端口	$S_{下}$	25G
	叶交换机	上联端口	$L_{上}$	25G
		下联端口	$L_{下}$	10G
设计参数	$L_{带宽上}/L_{带宽下} \geq 1/3$			
设计能力	可支持服务器数量	$S_{下} \times L_{下}$		
	北向总带宽	$L_{上} \times 25G$		
	南向总带宽	$L_{下} \times 10G$		

注:①脊交换机下行端口数量,决定了叶交换机的数量,本例中叶交换机有 $S_{下}$ 台。②叶交换机上行端口数量,决定了脊交换机的数量,本例中脊交换机有 $L_{上}$ 台。③脊交换机下行端口数量和叶交换机上行端口数量共同决定了叶脊网络的规模。

目前,主流厂商已经提供一系列的叶交换机和脊交换机,为有需要的数据中心用户搭建叶脊网络架构提供设备支持,表4示出的是些相关产品参数。

表4 40/100G叶脊交换机产品

产品名称	产品性能参数
叶交换机	交换机容量:800 Gbit/s~1.44 Tbit/s
	非阻塞带宽接入:400~720 Gbit/s
	虚拟化支持:MLAG、VxLAN
	端口配置1:32×10G+3×40G
	端口配置2:48×10G+4×40G
脊交换机	交换机容量:1.92~2.4 Tbit/s
	非阻塞带宽接入:960 Gbit/s~1.2 Tbit/s
	虚拟化支持:MLAG、VxLAN
	端口配置1:20×40G+4×100G
	端口配置2:48×10G+2×40G+4×100G

数据中心服务器通过VMware、KVM、Xen、Virtual Box等Hypervisor工具虚拟化之后,东西流量将达到总带宽的77%,跨数据中心为9%,南北流量仅占总带宽的14%。因此,还需要OpenStack这类虚拟化管理平台,在大二层的网络连接中,将网络、计算、存储和安全功能全部虚拟化,融合在一套虚拟化管理软件平台中,对用户提供虚拟存储、网络、计算资源等友好的交互界

面。

另外,虚拟化技术也可以根据实际情况采用 Docker 技术,配合 K8S、Compose、Marathon、Swarm、Mesos 等容器管理平台,达到上述“超融合”平台。

3 叶脊架构对数据中心带来的挑战

叶脊网络架构允许终端之间以最高万兆每秒的速度在任意的通信模式下进行通信,这种新型“胖树”能够很好应对数据中心激增的数据量(尤其是东西向流量),满足数据中心内部高速互连的需求。但这种扁平化的大二层叶脊架构同时给数据中心带来了新的挑战。

首先,叶脊网络架构对光模块的需求将大幅增加,相比传统三层硬件架构,差别最高可达到 30 倍之多(见表 5)。

表 5 2 种架构下光模块对比

数据中心类型	架构	光模块需求		
		10G	40G	100G
中小型 (1 000 台服务器)	传统架构	2 000	16	4
	叶脊架构	1 920	160	16
	差别(叶脊架构/传统架构)	0.96	10	4
大型 (1 000 个机柜)	传统架构	128 000	160	8
	叶脊架构	120 000	4 800	32
	差别(叶脊架构/传统架构)	0.94	30	4

2 种架构下,数据中心光模块在 10G 光模块需求下差别不大,在 40G 或者 100G 光模块需求下出现明显差别,尤其是在 40G 光模块需求下,差别需求达到 30 倍。

其次,叶脊网络大二层架构如何在虚拟化的技术背景下,实现用户良好感知的虚机迁移。

再者,随着网络规模的增加,二层链路层数据报文也将明显增加,如何保障正常的网络流量,也是需要面对的新挑战。

4 结束语

虽然传统的三层网络架构在很长一段时间内支撑了各种类型的数据中心,但是随着业务和技术的发展,90% 以上云数据中心流量和计算任务会通过数据中心进行处理。基于南北流量设计的传统三层硬件架构已经不能满足业务发展趋势,企业将面临成本和可扩展性的战略困境。

新的叶脊网络采用扁平化的大二层架构,扩大了

接入和汇聚层,大大提高了网络效率,而且不需要采购高性能的核心交换机,由相对较小规模的中低端交换机组网,以 40G/100G 带宽部署为主(大型数据中心可以考虑前瞻性部署 200/400G 带宽),支持全速东西向流量,同时可以减少企业部署和维护的费用,为数据中心提供一种切实有效的发展战略。

参考文献:

- [1] 2019-2025 年中国互联网数据中心(IDC)IT 行业深度分析及发展前景预测报告[EB/OL]. [2020-03-12]. <https://wenku.baidu.com/view/dc96360670fe910ef12d2af90242a8956aecaad2.html>.
- [2] 数据中心白皮书(2018)[EB/OL]. [2020-03-12]. http://www.caict.ac.cn/kxyj/qwfb/bps/201810/t20181016_186900.htm.
- [3] 2020-2025 年中国互联网数据中心行业市场前景及投资机会研究报告[EB/OL]. [2020-03-12]. <https://wenku.baidu.com/view/52785e2118e8b8f67e1cfad6195f312b3069eb98.html>.
- [4] 2019 年数据中心 IDC 行业深度研究报告[EB/OL]. [2020-03-12]. <http://www.199it.com/archives/981608.html>.
- [5] 三大需求拉动光纤光缆行业重迎增长机遇[EB/OL]. [2020-03-12]. https://finance.sina.com.cn/stock/hyyj/2020-07-10/doc-iirczym1594008.shtml?cre=tianyi&mod=pcpager_fin&loc=32&r=9&rfunc=100&tj=none&tr=9.
- [6] 光通信的再思考:5G 流量爆发下的数据密度革命[EB/OL]. [2020-03-12]. <https://baijiahao.baidu.com/s?id=1665549431284268486&wfr=spider&for=pc>.
- [7] IDC 叶脊架构如何推动高端光模块爆发[EB/OL]. [2020-03-12]. <https://xuanguobao.cn/article/128584>.
- [8] 数据中心 spine leaf 网络架构[EB/OL]. [2020-03-12]. <http://win-hong.com/news/detail.aspx?id=100000570089826>.
- [9] 余嘉历. 应用叶脊网络架构改造企业数据中心网络[J]. 数字化用户, 2019, 25(29): 68.
- [10] 山姆·哈拉比. 超融合基础架构(HCI)数据中心[M]. 田果, 译. 北京: 人民邮电出版社, 2020.
- [11] 郭亮. 数据中心热点技术剖析[M]. 北京: 人民邮电出版社, 2019.
- [12] 张磊, 陈乐. 国之重器出版工程 云数据中心网络架构与技术[M]. 北京: 人民邮电出版社, 2019.
- [13] 李劲. 云计算数据中心规划与设计[M]. 北京: 人民邮电出版社, 2018.
- [14] 张晨. 云数据中心网络与 SDN: 技术架构与实现[M]. 北京: 机械工业出版社, 2018.

作者简介:

包琅允, 毕业于湘潭大学, 高级工程师, 学士, 主要从事核心网, 云计算和大数据等工作。

