

跨DC的虚拟化核心网

Research on the Disaster Recovery System of
Cross-DC Virtual Core Network

容灾体系研究

韦国锐, 陈立栋, 于秋思, 杨 晓 (中国联通广东分公司, 广东 广州 510000)

Wei Guorui, Chen Lidong, Yu Qiusi, Yang Xiao (China Unicom Guangdong Branch, Guangzhou 510000, China)

摘 要:

为了实现“弹性、伸缩、快速业务迁移”的网络特性,基于以下原则重新设计NFV系统:前端无状态、横向可扩展、单点失效不影响业务,分离业务处理单元和存储节点。跨DC的虚拟化核心网容灾体系可以利用现有技术,在不同的通信云之间近乎无损地迁移虚拟化网元。

关键词:

虚拟化;核心网;跨DC;容灾体系;业务迁移
doi: 10.12045/j.issn.1007-3043.2019.09.017
中图分类号: TN915
文献标识码: A
文章编号: 1007-3043(2019)09-0078-04

Abstract:

To achieve the goal of "flexible, scalable and fast business migration", the NFV system is redesigned based on the following principles: front-end statelessness, horizontal scalability, single point of failure does not affect the business. Business processing units and storage nodes are separated. Disaster recovery system of cross-DC virtual core network can use existing technology to migrate virtualization network elements almost non-destructively between different communication clouds.

Keywords:

Virtualization; NFV; Core network; Cross DC; Disaster recovery system

引用格式: 韦国锐, 陈立栋, 于秋思, 等. 跨DC的虚拟化核心网容灾体系研究[J]. 邮电设计技术, 2019(9): 78-81.

1 概述

近年来,基于自建云网络部署虚拟化电信核心网已经成为当前进行4G核心网以及后续5G核心网部署的必然选择。2015年2月中国移动提出了NovoNet计划作为其未来网络演进的基础,2015年9月中国联通提出了基于SDN、NFV和云技术的网络重构基础架构CubeNet2.0,2016年6月中国电信提出了CTNet2025,为未来10年网络重构制定了相关规划。各家电信运营商的网络重构计划都是基于NFV和云化网络。以CUBE-Net 2.0为例,其核心技术是SDN、NFV和Cloud,

它以超宽网络和数据中心为载体,通过云化的服务平面,实现网络集约化运营,适用于多种应用服务场景。

从3GPP协议的发展来看,移动网络从Rel13的4G核心网,演进到Rel14的EPC转控分离,再到Rel15的5G NAS体系架构,最终演进到Rel16的5G SA体系架构,从Rel14转控分离开始,移动网络基本都是以通信云(DC)为基础,而目前可选的灾备方案中,池组化灾备已经成为大部分运营商及厂家的选择。以某厂家在通信云上的虚拟化移动分组网为例,采用了2套MME组成一个容灾池组,实现多DC间MME的主备容灾部署。

池组化容灾模式是在2个不同DC上分别部署MME网元,一个为主网元,另一个是备网元(冷备方式

收稿日期: 2019-06-10

工作)。主备MME对周边网元发布高低优先级不同的相同业务路由,由于同一时刻只有发布高优先级路由的网元可以和周边网元互通,低优先级路由不会进入路由表,且主备MME业务逻辑地址相同,因此相对于周边网元主备MME整体呈现为一个逻辑网元。

2个DC间的MME通过协商确定各自的主备身份,并基于多通道(容灾控制通道、容灾业务通道)的检测结果进行故障仲裁,主网元故障后,路由收敛到备网元,业务在备网元上重新接入。

在虚拟机出现故障时,虽然上述容灾方案可以将用户业务倒换至冷备DC中的虚拟核心网网元上,但还是存在以下问题。

a) 用户业务被中断后会不断尝试在冷备核心网网元中重新建立连接状态。

b) 冷备DC中也需要建立一套虚拟化核心网网元,并保持对外连接,在计算资源、存储资源、链路资源、动力等方面存在一定的浪费。

c) 在5G时代,随着网络切片技术的应用普及,同一个通信云内的VNF网元数量会增多,网络环境变得更加复杂,每个虚拟网元承载的业务量会减少,同时当前标准定义的eMBB、uRLLC、mIoT 3类切片将承载在不同的虚拟网元上,与现有网络承载模式不同。如果保持现有的池组化容灾体系,会增加网络配置复杂度,降低管理效率。

在NFV架构下,如果计算节点同时存储用户的状态数据(即用户上下文信息),那么在删除计算节点之前需要将用户上下文信息(包含IMSI、APN、QoS、速率、PDP类型、PDP地址等)迁移到其他节点上,否则会造成业务损失;同时在新增计算节点时需要将用户上下文信息迁移过去,否则新增节点将不具备处理在线用户数据业务的功能。

2 跨DC容灾系统设计

为了实现“弹性、伸缩、业务快速迁移”的网络特性,基于以下原则重新设计NFV系统:前端无状态、横向可扩展、单点失效不影响业务,分离业务处理单元和存储节点。计算节点不再保存用户状态信息,只负责对信令进行处理。业务处理单元在处理用户消息(信令)时需要从数据库中获取用户信息,再结合用户的状态对消息进行处理。这也就是将一个传统网元拆成多个虚拟网络功能组件(VNFC——Virtualized Network FunctionComponent),具体的业务处理结构如图1所示。

图1所示。

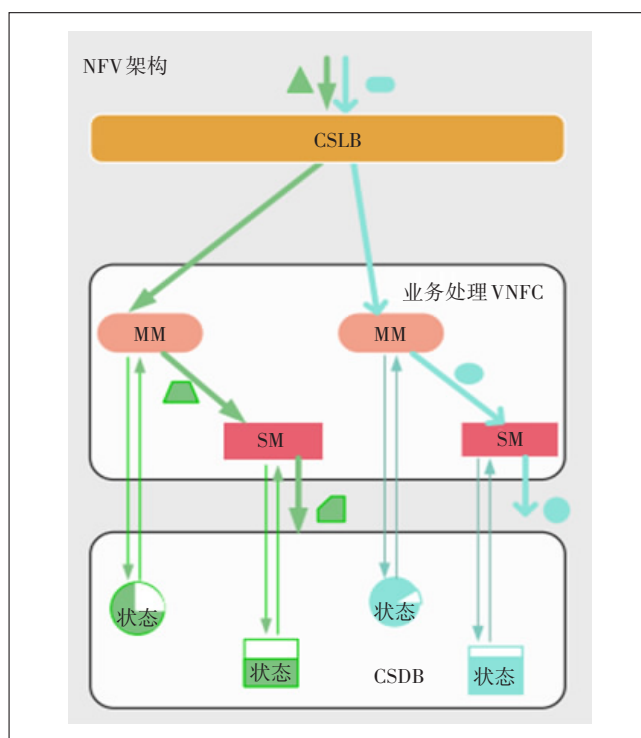


图1 业务处理结构

基于上述业务处理结构,虚拟机可以在网络中不受限迁移,这也就实现了业务的灵活变更。本文采用图2的方案对移动核心网虚拟网元进行业务无损容灾迁移,方案采用VxLAN和EVPN技术对虚拟网元中的计算资源进行大二层容灾迁移;并基于原有SAN存储模式的异地容灾系统,采用实时异步复制数据模式,在同一个地(市)的不同DC间进行容灾。

2.1 虚拟机(VM)的跨DC迁移

虚拟机迁移,顾名思义,就是将虚拟机从一个物理机迁移到另一个物理机,但是要求在迁移过程中业务不能中断。要做到这一点,需要保证虚拟机迁移前后,其IP地址、MAC地址等参数维持不变。这就决定了,虚拟机迁移必须发生在一个二层域中。所以虚拟机的跨DC迁移就需要在DC之间建立一个大二层的连接通道,基于现有技术和可维护性,本文采用VxLAN和EVPN技术在DC间建立一个大二层连接通道,以实现虚拟机的跨DC迁移。

2.1.1 VxLAN和EVPN技术

VxLAN是由IETF定义的NVO3(Network Virtualization over Layer 3)标准技术之一,采用L2 over L4(MAC-in-UDP)的报文封装模式,将二层报文用三层

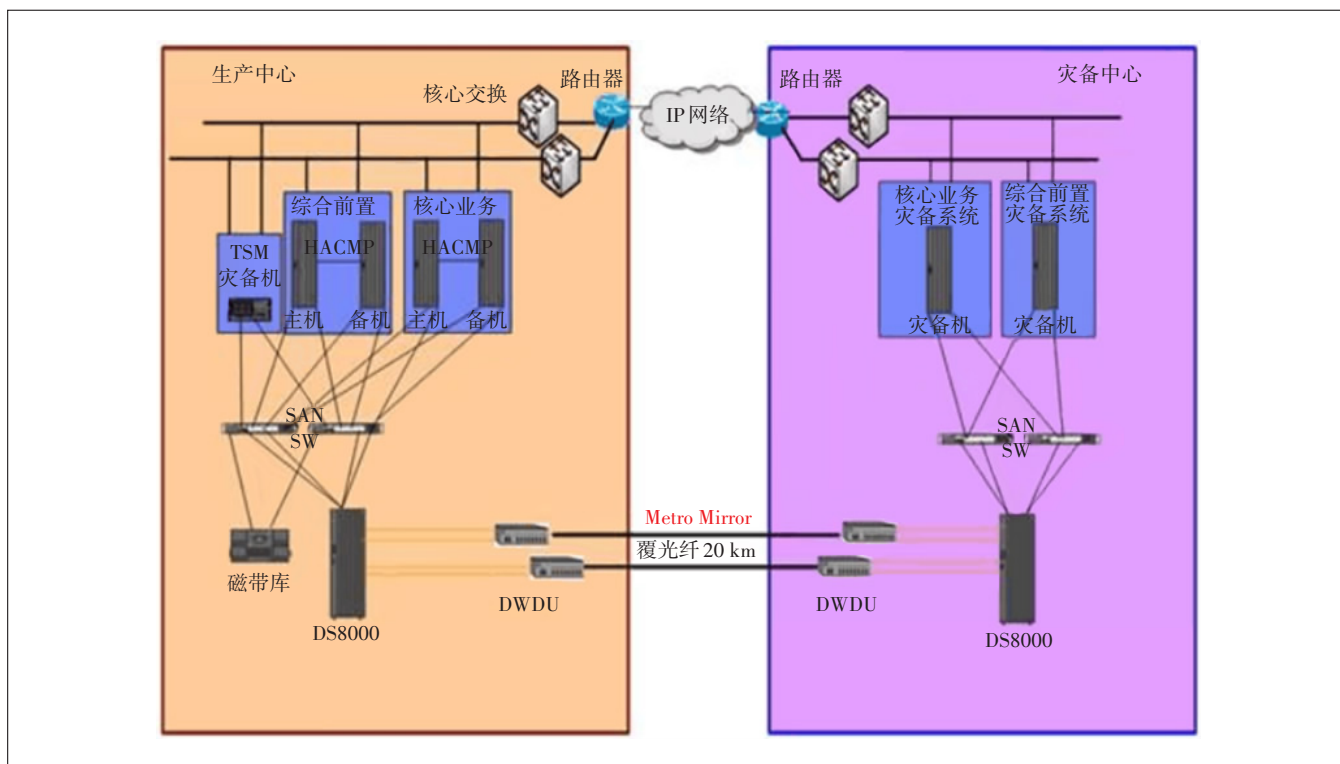


图2 业务无损容灾迁移网络结构

协议进行封装,可实现二层网络在三层范围内进行扩展,同时满足数据中心大二层虚拟迁移和多租户的需求。

最初的VxLAN方案(RFC7348)中没有定义控制平面,是通过手工配置VxLAN隧道,然后用流量泛洪的方式进行主机地址的学习。这种方式实现上较为简单,但是会导致网络中存在很多泛洪流量,网络扩展的难度较大。

为了解决上述问题,VxLAN引入了EVPN(Ethernet VPN)作为其控制平面。EVPN参考了BGP/MPLS IP VPN的机制,通过扩展BGP协议定义了几种BGP EVPN路由,通过在网络中发布路由来实现VTEP的自动发现、主机地址学习。

采用EVPN作为控制平面具有以下一些优势。

- a) 可实现VTEP自动发现、VxLAN隧道自动建立,从而降低网络部署、扩展的难度。
- b) EVPN可以同时发布二层MAC和三层路由信息。
- c) 可以减少网络中泛洪流量。

2.1.2 虚拟机跨DC迁移的网络设计

大二层的DCI连接建立时,VM1和VM2间的通信采用三段式VxLAN的方案,在数据中心A和数据中心

B内部分别创建VxLAN隧道,在数据中心的出口路由器(Transit Leaf)间也创建VxLAN隧道。同时在数据中心内部和跨数据中心互联链路上分别配置BGPE-VPN,具体如下。

a) 控制平面:Server Leaf 1在学习到VM1的MAC地址后,生成BGP EVPN路由发送给出口路由器1(Transit Leaf 1);出口路由器1收到BGP EVPN路由后,先交叉到本地EVPN实例中,并在EVPN实例中生成VM1的MAC表项,再通过BGP EVPN方式将VM1的MAC地址更新发送给出口路由器2(Transit Leaf 2);出口路由器2再广播至Server Leaf 2,在Server Leaf 2中形成VM1的MAC地址项。同时,VM2的MAC地址经由相同的方式转发到Server Leaf 1,这样VM1和VM2之间就建立了可以直接进行二层数据转发的路径。

b) 数据转发平面:通过图3所示的三段式VxLAN隧道,系统利用控制平面生成的MAC地址转发表,将数据逐段转发至对端VM。在Transit Leaf 1和Transit Leaf 2间建立端到端的VxLAN隧道时,需要实现Vx-LAN Mapping功能,并进行一次VNI的转换。

2.2 基于SAN存储模式的异地容灾系统设计

虚拟机通过存储区域网络(SAN——Storage Area

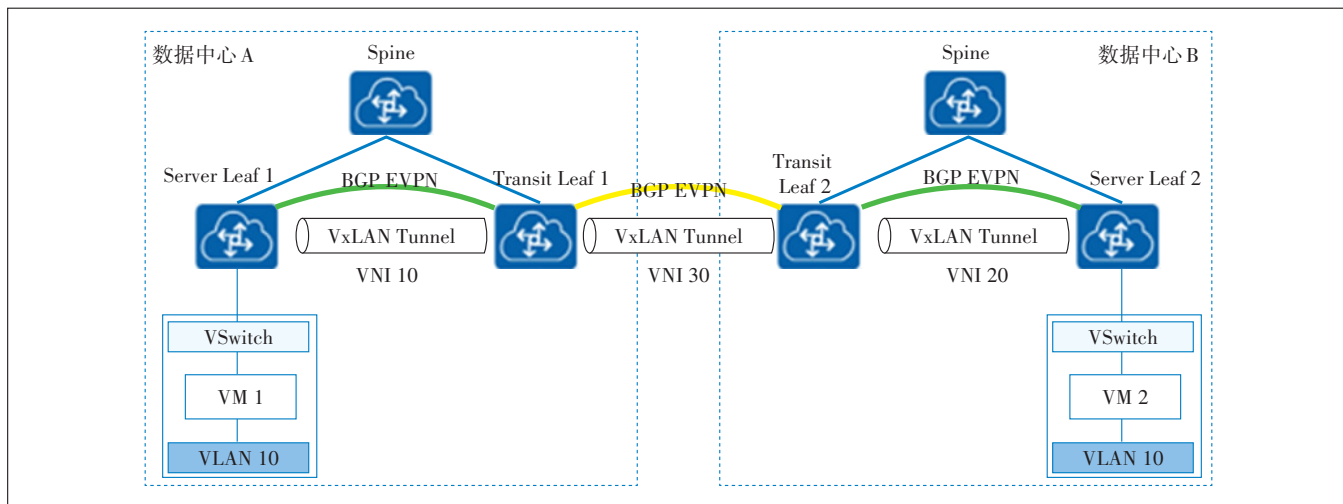


图3 数据转发平面

Network)与存储设备相连。随着以太网技术的发展,特别是10GE网络的普及,以太网的性能已经能满足存储的要求。当虚拟机有数据存取的需求时,数据可以通过IP SAN网络高速传输。基于SAN存储模式的异地容灾系统如图4所示,其采用异步复制的方式同步主备数据中心的数据。各个设备厂家均有成熟的方案,在此不再赘述。

需求,而且在主备网元切换过程中用户业务必须中断。而本文中的跨DC的虚拟化核心网容灾体系可以利用现有技术,在不同的通信云之间,近乎无损地迁移虚拟化网元,新的容灾体系能更好地适应通信云中虚拟机常态化迁移的特点,增加了数据中心的计算密度,可以按需启用数据中心服务器和存储资源,减少了人工管理的成本,提升了用户业务感知。

参考文献:

- [1] 张晨. 云数据中心网络与SDN[M]. 北京:机械工业出版社,2018.
- [2] C FILSFILS. Segment Routing 详解(第一卷)[M]. 北京:人民邮电出版社,2017.
- [3] 朱常波. 电信云的发展和对新业务的支撑思考[J]. 邮电设计技术,2018(1):33-37.
- [4] 陆钢,杨新章,何震苇,等. 电信运营商对容器技术的探索和思考[J]. 电信科学,2016,32(8):159-163.
- [5] 张艺林. 运营商交换核心网元容灾备份技术的应用[J]. 电子技术与软件工程,2017(9):188-188.
- [6] 刘晓燕,郑慧泳. 利用CLUSTER技术实现MGW自动容灾[J]. 信息通信,2018,186(6):72-75.
- [7] 陆广海. 爱立信核心网CSFB快速容灾方法研究[J]. 电信快报:网络与通信,2018(2).
- [8] 刘海涛,王雪涛,马燕茹. 移动核心网虚拟化容灾方案研究[J]. 电信快报,2017(9).

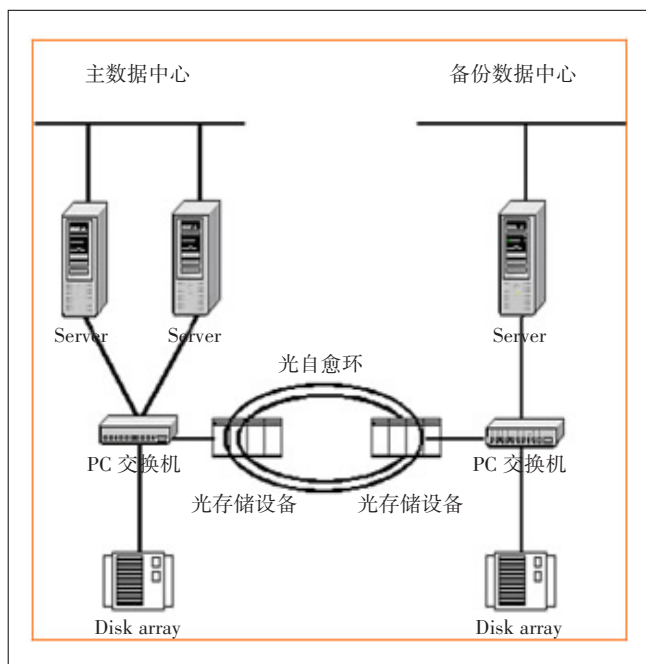


图4 容灾备份结构

3 结束语

传统的容灾方案已经无法适应不同业务的质量

作者简介:

韦国锐,毕业于华南理工大学,硕士,主要从事移动通信网络维护优化工作;陈立栋,毕业于北京邮电大学,硕士,主要从事移动通信网络维护优化工作;于秋思,毕业于西安电子科技大学,学士,主要从事移动通信网络维护优化工作;杨晓,毕业于华南师范大学,硕士,主要从事移动通信网络维护优化工作。