ZTE vDC 网络解决方案技术白皮书

目录

1 技术背景

过去的二十年间，随着互联网的发展，越来越多的互联网业务以数据中心的方式提供，大型企业也逐步建立自己的数据中心来承载自营IT系统。由于数据中心的规模效应显著，单位成本随着规模的上升而急剧下降，尤其表现在场地、供电、散热、运维、出口带宽等方面的成本，因此承载在大型数据中心之上的云计算技术会成为主流的应用。

从技术的角度来看，按需提供计算基础设施、服务即可以称之为云，包括计算、存储、网络资源的管理调度系统，各种供客户调用的软件服务接口、以及多租户的SaaS系统。云计算发展的早期致力于计算资源的虚拟化和按需分配管理。随着云计算技术成熟度逐渐完善，云计算技术在提高资源利用率、推进工作方式变革等方面的优势也凸显出来，越来越多的用户选择采用云计算技术构建新型数据中心。

- 数据中心已经成为电信运营商业务容器和核心载体，通过建设新型数据中心以承载各类NFV云化软件及IT系统，实现ICT融合。
- 政企用户通过建设新型数据中心，企业工作在云端，通过云数据中心的整体分布式计算能力来处理工作，变革企业工作方式，提升运作效率。
- 互联网企业提供公有云服务，用户像使用水电等公共服务一样的使用计算资源。

相应地，云计算技术对IDC网络也提出了新的要求，即IDC网络需提供网络资源虚拟化，实现“云”和“网络”资源的统一规划部署和调度，以满足云数据中心的建设需求。因此vDC网络技术也应运而生。

vDC网络是指在已有的物理网络上虚拟出来客户化的网络组件，自动化、自服务，让客户感觉和独占的物理网络体验相似。vDC网络技术需要提供的功能包括：

1. 网络连接通信功能

- 叠加网络（Overlay）功能

叠加网络（Overlay）是一种网络架构上叠加的虚拟化技术模式，在对基础网络不进行大规模修改的条件下，实现应用在网络上的承载，并可以实现Overlay叠加网络业务与Underlay基础网络业务分离。

- 租户（Tenant）隔离功能

租户是云计算环境中管理的主体，vDC网络技术以租户为单位进行资源池的应用和互通，租户之间的网络、路由相互隔离。

- 虚拟网络（Network）隔离功能

虚拟网络是一种逻辑的抽象网络，在vDC网络技术中，通过虚拟网络实现网络的隔离。相同虚拟网络之内的主机（VM或裸金属服务器）可以互通，而不同虚拟网络之间的主机（VM或裸金属服务器）无法互通。（注：可以通过路由器或外部网络互通）

- 虚拟网络子网（Subnet）功能

虚拟网络子网是虚拟网络的一部分，标识了承载在虚拟网络上的网络段。vDC网络技术可以利用虚拟网络子网承载子网网段，DHCP地址池，网关，DNS Server地址等资源。

- 虚拟机接入虚拟网络功能

在云计算资源池中，通过计算资源虚拟化提供虚拟机业务，vDC网络技术需要为虚拟机提供网络服务，使虚拟机可以接入到虚拟网络，实现虚拟机与虚拟机，虚拟机与外部网络之间的互通。

- 裸金属服务器接入虚拟网络功能

除了虚拟机业务外，云计算资源池中也包括了非虚拟化的裸金属服务器等计算资源，vDC网络技术需要为裸金属服务器提供网络服务，使裸金属服务器可以接入到虚拟网络，实现裸金属服务器与虚拟机，裸金属服务器与其他裸金属服务器，裸金属服务器与外部网络之间的互通。

- 虚拟路由器功能

虚拟路由器负责同一租户内的虚拟网络之间的互联，虚拟路由器提供虚机接口接入到虚拟网络，提供虚拟网络子网的网关服务，实现不同子网之间的三层转发。

- 虚拟网络连接外部网络功能

云计算资源池中的虚拟机或裸金属服务器除了需要与内部的主机进行互通外，还需要与vDC网络之外的网络进行互联。因此vDC网络技术需要为虚拟机或裸金属服务器提供外部网络互联的能力，实现内部主机与外部网络的互通。

- 浮动IP/NAT功能

云计算资源池中的虚拟机或裸金属服务器在与外部网络互联时，需要使用外部网络的IP地址空间，比如使用公网的IP地址编址，而vDC网络通常采用私有地址池构建，因此当vDC网络内的虚拟网络与外部网络互通时，需要vDC网络技术为虚拟网络提供地址转换服务。

浮动IP，也就是1:1NAT服务，即需要为每一个内部的虚拟机或裸金属服务器的私网IP分配一个外网的IP地址，通常应用在资源池提供Server服务，外部主机主动访问vDC网络内的主机的情况。

N：1 NAT服务，即为多个内部的虚拟机或裸金属服务器的私网IP分配一个外网的IP地址，通常应用在vDC网络内的主机主动访问外部网络资源的情况。

2. 网络管理功能

- 网络地址管理功能

vDC网络中的虚拟机需要通过DHCP方式获取IP地址，因此需要vDC网络技术提供网络地址管理功能，提供DHCP Server服务，并分配网关，DNS Server地址等资源，实现虚拟机的IP地址管理。

- VNI管理功能

vDC网络基于VxLAN提供叠加网络服务，同时通过VNI实现虚拟网络的隔离，这就要求了vDC网络技术能提供VNI的资源管理服务。

- 访问控制策略管理功能

从安全的业务考虑，vDC网络内的虚拟机需要通过访问控制策略，进行访问控制，实现虚拟机资源的保护。vDC网络技术通过访问控制策略实现虚拟机之间的访问控制。

- QoS管理功能

vDC网络技术要求支持QoS管理功能，支持DSCP重标记，流量速率限制等网络服务。

- 虚拟机位置迁移功能

vDC网络中的虚拟机资源，可以进行位置的迁移，这就需要vDC网络技术能为虚拟机位置迁移提供服务，迁移前后的网络服务、访问控制策略、QoS策略均需要保持不变。

3. 运维管理功能

- 虚拟拓扑呈现

vDC网络中包括了虚拟网络，虚拟路由器等逻辑网络资源，为了便于vDC网络的统一管理，vDC网络技术需要支持虚拟拓扑的呈现，展示虚拟网络层次的拓扑管理。

- 物理拓扑呈现

vDC网络通过叠加网络技术实现应用在网络上的承载，因此虚拟网络上的维护管理，需要关联物理网络的拓扑。

- 流量监控

vDC网络技术要求支持流量监控功能，对虚拟机端口、VxLAN隧道、特定流进行流量监控。

- 流量告警

vDC网络技术要求支持流量告警功能，对流量监控设置告警阈值，当流量超过阈值时能进行流量告警。

- 告警管理

vDC网络技术要求支持告警管理功能，对流量、资源使用情况、网络性能数据进行阈值设置，对超过阈值的事件进行告警。

- 日志管理

vDC网络技术要求支持日志管理功能，用户操作，网络事件，系统日志进行记录。

- 流量镜像

vDC网络技术要求支持流量端口镜像，能够对指定端口的流量进行镜像。

vDC网络技术要求支持流量流镜像，能够定义流规则，并对匹配流规则的流量进行镜像。

- 故障诊断

vDC网络通过叠加网络技术实现应用在网络上的承载，这就给故障诊断带来新的挑战。为了进行问题的定位，需要vDC网络技术提供端到端的故障诊断，满足Overlay、Underlay等不同层次的故障诊断要求。

- 性能统计

vDC网络技术要求支持性能统计，能够对虚拟网络、子网、虚拟路由器等虚拟网络对象，以及Underlay层次的控制器、NVE、网关、虚拟防火墙、虚拟负载均衡器等网络对象的关键指标进行性能统计。

4. 增值服务功能

- 防火墙功能

vDC网络中的虚拟机或裸金属服务器与外部网络互通时，需要采用防火墙进行流量的安全防护。为适应vDC网络的多租户管理模型，vDC网络技术需要为每个租户提供不同的防火墙服务，每个租户的防火墙部署不同的安全策略。

- 负载均衡器功能

vDC网络中的虚拟机或裸金属服务器提供L4-L7层服务时，考虑到性能和可靠性因素，需要构建多虚拟机集群提供服务，并采用负载均衡器保障服务在多虚拟机集群中进行负载分担。为适应vDC网络的多租户管理模型，vDC网络技术需要为每个租户提供不同的负载均衡器服务，每个租户的负载均衡器部署不同的负载分担策略。

- IPSec VPN功能

vDC网络中的虚拟机或裸金属服务器需要与企业私有云或园区网络互通时，需要采用IPSec VPN提供安全隧道，保障通信的私密性。为适应vDC网络的多租户管理模型，vDC网络技术需要为每个租户提供不同的IPSec VPN服务，每个租户的IPSec VPN服务部署不同的访问策略。

2 vDC 网络技术优势

基于在数据中心建设和数据通信方面的技术积累，中兴通讯通过vDC网络技术提供开放、灵活的能力，可助力运营商、政企客户构建新型的云数据中心网络。

vDC网络技术具有以下技术优势：

1. Overlay网络，在传统IP网络上构建SDN Fabric

Overlay是一种网络架构上叠加的虚拟化技术模式，其可以在对基础网络不进行大规模修改的条件下，实现应用在网络上的承载。vDC网络技术通过Overlay网络来构建数据中心基础网络架构，具备了以下优点：

- 利用VxLAN构建Overlay网络，承载网络无特殊拓扑限制，IP可达即可；对现网络改动较小，保护用户现投资。
- 支持16M多租户/网络，极大扩展了隔离数量。
- 通过VxLAN承载大二层网络，承载网络无需大量VLAN Trunk，简化了网络结构，避免了环路、广播风暴的风险。
- 支持虚拟机灵活迁移，网络策略动态跟随。
- 通过主机/ARP代理学习，消除了广播、未知单播的泛滥。
- 通过SDN进行隧道管理，转发控制，可以实现分布式路由转发，优化转发路径。

2. 可视化管理，优化网络调度

vDC网络技术通过集中式的SDN控制器，提供设备接入管理，网络拓扑自动发现，转发路径计算，故障诊断，流量工程等可视化管理，优化网络的调度。

- 资源情况一目了然，一切尽在掌握
    - 用图形化界面掌握VLAN、物理端口、IP地址、MAC地址等资源使用情况，一目了然。
- 多角度实时流量监控

- 多种实时流量监视视图
- 基于历史流量数据生成基线
- 实时流量信息与基线对比，流量异常自动触发告警

- 全方位网络监控，端到端故障诊断
  - 提供系统资源查看，流表数量监控，流量监控，日志，告警，交换机及资源管理等
  - 控制器、转发面设备、业务等全面的告警。

3. 云管理平台集成，自动化部署网络服务

SDN控制器支持基于REST/RESTCONF的开放API接口。用户可以通过云管理平台，比如OpenStack，创建虚拟租户、网络资源，而OpenStack平台利用Neutron Plugin调用ZENIC控制器的北向API接口，创建/维护租户、网络、子网、端口等网络资源对象，由ZENIC vDC控制器实现网络资源对象的自动化部署，部署过程中无需人工参与。

vDC网络解决方案，可以支持异构云管理平台、异构Hypervisor的资源池管理。

4. 业务链，定制化编排网络服务

传统IDC中，增值服务在网络部署阶段已经决定了服务的顺序和种类，无法满足用户自定义网络服务的需求。而在虚拟化数据中心中，租户会根据实际的业务模型，动态选择服务类型（比如FW，LB，VPN等），并对服务的执行顺序进行编排。中兴通讯vDC网络解决方案，利用业务链技术，定制化编排网络服务，满足用户的个性化需求。

3 背景技术

3.1 SDN 技术

SDN技术起源于斯坦福大学Clean Slate项目所提出的OpenFlow技术，并在2011年ONF成立后进一步扩大了业界的影响。在早期的SDN发展中，业内形成了两种不同的技术路线：

- ONF所倡导的基于纯OpenFlow以及后续演进的P4 Runtime的完全控制转发分离技术。该技术流派的优点是转发面简单化、白盒化，智能集中在控制器。此种方案强化了SDN控制器的地位，降低了对转发设备的要求，也降低了转发设备的价值。

- 业界部分厂商所采取的演进式SDN方案，侧重于网络的自动化，网络设备保留完全的智能，提供API给控制器进行网络的自动化编排，比如IETF标准化的EVPN分布式控制面+NetConf配置自动化接口，以及在WAN网络中基于PCEP、NetConf、BGP-FlowSpec控制接口的SDN方案。

中兴通讯vDC网络技术采用Overlay技术，OpenFlow控制vSwitch，对于硬件交换机网络采用BGP EVPN组成的完全控制转发分离技术来构建vDC网络，同时也支持vSwitch和硬件交换机构成的混合Overlay。

3.2 Overlay 技术

Overlay是一种网络架构上叠加的虚拟化技术模式，其可以在对基础网络不进行大规模修改的条件下，实现应用在网络上的承载。

VxLAN(virtual Extensible LAN)虚拟可扩展局域网，是一种Overlay的网络技术，使用MAC in UDP的方法进行封装。VxLAN提供了将二层网络Overlay在三层网络上的能力，VxLAN Header中的VNI有24个bit，数量远远大于4096，并且UDP的封装可以穿越三层网络，比VLAN有更好的扩展性。

VTEP 是指VxLAN隧道终端 (VxLAN Tunneling End Point)，用于对VxLAN报文进行封装/解封装。在一端VTEP封装报文后通过隧道向另一端VTEP发送封装报文，另一端VTEP接收到封装的报文解封装后，根据封装的MAC地址进行转发。

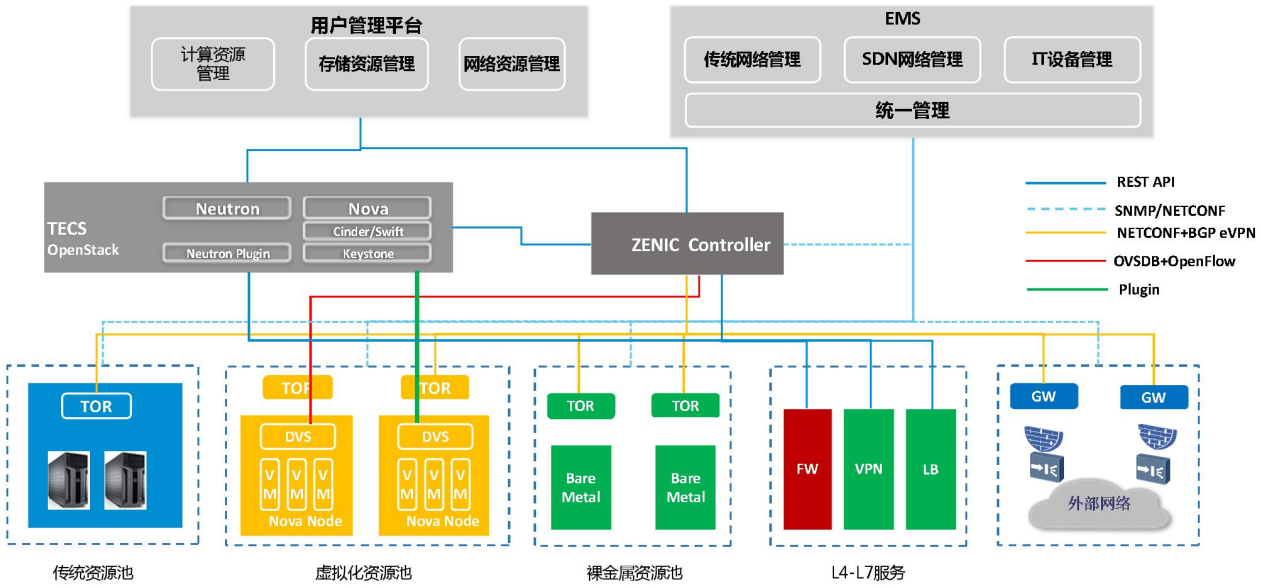
VTEP设备按形态通常分软件VxLAN隧道终点和 硬件VxLAN隧道终点。软件VxLAN隧道终点，比如开源软件OpenvSwitch等，可以提供VxLAN隧道封装，并可以支持基于OpenFlow的流表转发能力。硬件VxLAN隧道终点，通常用于提供VxLAN-VLAN转换能力，并且支持基于OpenFlow的转发管理，实现VxLAN网络与传统网络的互通。

中兴通讯vDC网络技术采用基于VxLAN的Overlay技术构建数据中心基础网络架构。

4 vDC 网络架构

4.1 vDC 网络逻辑架构

vDC网络采用SDN技术，通过转发控制分离、集中管控、开放可编程的网络体系架构，构建数据中心网络。



vDC网络逻辑架构中，包括的组件有：

- ZENIC vDC 控制器

ZENIC vDC控制器聚焦云数据中心的SDN需求，定位于私有云、公有云、NFVI和混合云的应用场景，结合业务编排器和云平台，提供电信级vDC网络的端到端SDN解决方案。

- EMS

EMS 网管是一套WEB化的网络管理系统，实现了物理网络资源、SDN控制器、虚拟网络资源统一管理和集中的操作维护。

- 云管理平台层

云管理平台层是提供计算、存储、网络统一管理的系统，常见的云管理平台包括OpenStack，CloudStack等，vDC网络技术需要提供云管理平台插件，为云管理系统提供网络实现。

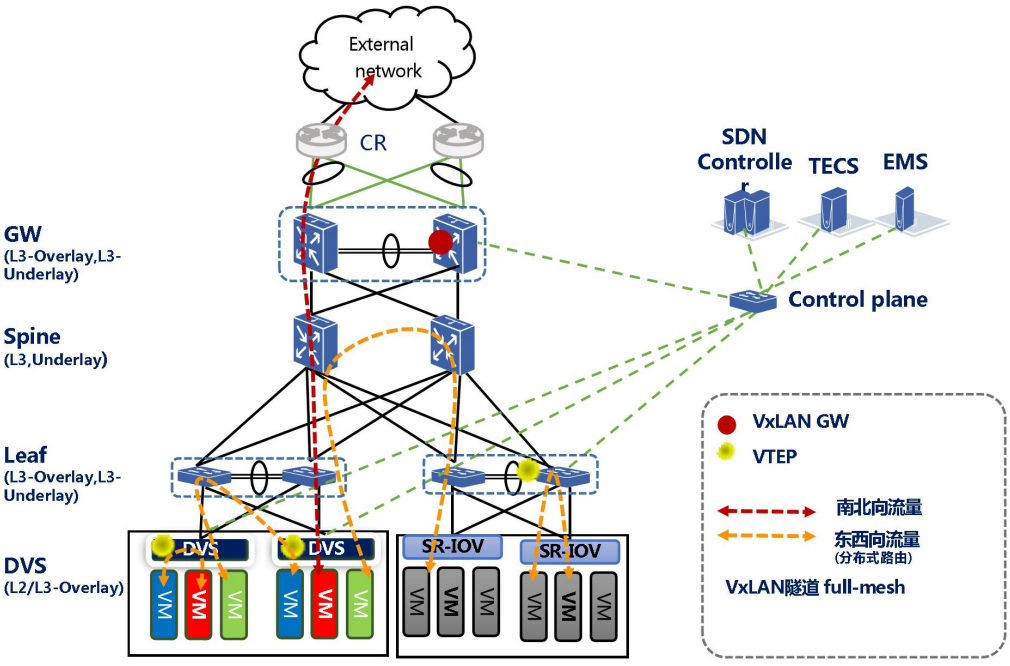
- 转发设备层

转发设备层是提供网络转发能力的实体，包括了运行在计算服务器的vSwitch软件设备，运行在机架的TOR硬件设备，负责VxLAN网络与外部网络互通的VxLAN GW设备，以及负责提供网络功能的软件L4-L7设备或硬件L4-L7设备等。

4.2 vDC 网络基础组网架构

4.2.1 混合 Overlay 组网

中兴vDC网络解决方案通过VTEP和VxLAN网关之间建立VxLAN隧道，来实现Overlay网络端到端打通，Overlay网络支持多租户、大二层以及按需配置和自动化部署。在混合Overlay组网方案下，VTEP既可以建立在硬件交换机上，也可以建立在虚拟交换机上。通常情况下，普通虚机采用vSwitch作为VTEP，SR-IOV虚机和裸金属服务器采用TOR交换机作为VTEP。



对于vSwitch做VTEP，控制平面基于OVSDB和OpenFlow，通过OpenFlow TTP表模型实现表项的控制和主机位置学习。

对于硬件交换机做VTEP(如：LEAF, GW)，控制面基于EVPN+NetConf，控制器通过NetConf配置ToR和GW的EVPN实例，VTEP间通过感知VRF内的主机上线情况来自动创建动态VxLAN Tunnel，转发面通过EVPN的RT-2和RT-5实现Server的MAC、ARP以及Subnet子网信息的学习。

通过控制器和网关之间的EVPN控制信道，网关可以将TOR交换机下挂的裸金属服务器、SR-IOV虚机路由信息上送控制器，用于生成OpenFlow流表下发给DVS；控制器也可以将vSwitch虚机信息通过eVPN控制通道，向硬件GW、硬件TOR进行重分发，从而实现整网的转发控制。

混合Overlay解决方案为满足数据中心特点，提供如下特色功能：

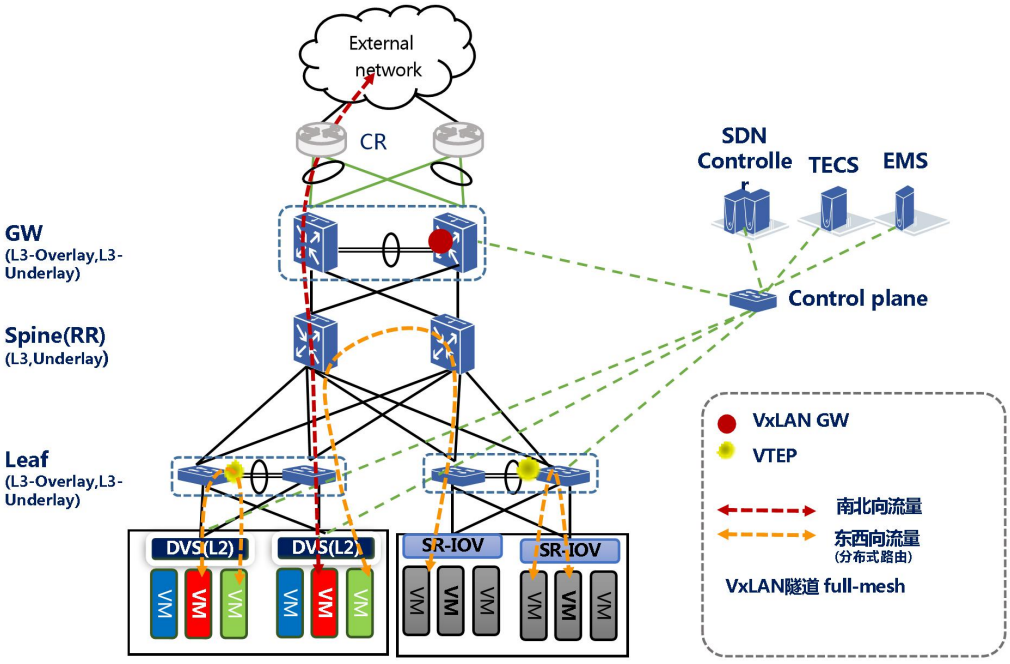
- 控制器通过虚拟L3默认网关的IP和MAC地址实现Anycast GW功能，无论VTEP是vSwitch还是ToR交换机，都支持跨子网的就近转发，降低通信延迟、降低GW负载。
- 开启ARP代答和主机ARP探测加速业务收敛，ARP和ICMP功能卸载到转发面，降低控制器CPU开销。
- 虚拟交换机ZXDVS基于Connection Track扩展实现有状态的L4轻量级分布式防火墙。
- 转发表项在交换机设备上保存，控制器离线后不会影响已建立服务通信。

4.2.2 硬件 Overlay 组网

在硬件Overlay组网方案中，BGP EVPN作为Overlay网络分布式控制面，在VTEP和VxLAN GW之间建立VxLAN隧道。无需对Underlay网络改造，构建Underlay基础上的Overlay网络，支持多租户、大二层以及按需配置和自动化部署。

硬件EVPN方案下，VTEP和VxLAN GW均采用硬件设备并将EVPN作为控制面协议。ZENIC vDC控制器负责自动化配置，控制器将云平台创建的router、network以及router interface对象分别映射为VRF、L2VNI以及L3接口，然后通过NetConf协议在相应交换机设备上配置。

硬件Overlay方案在只存在裸金属服务器的场景下部署相对简单。有些场景中部署vSwitch实现安全策略自动化和报文转发，推荐vSwitch工作在纯二层实现VM二层转发、安全策略和VM迁移策略跟随，通过硬件Overlay提供大二层网络和三层网关功能。

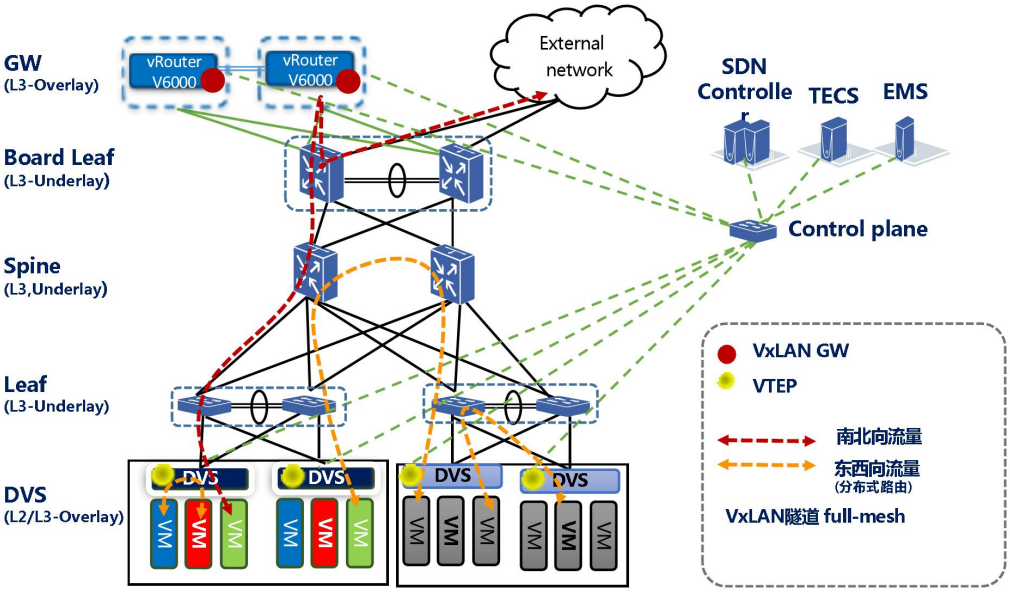


硬件 Overlay 解决方案为满足数据中心特点，提供如下特色功能：

- 在任意邻接的硬件交换机之间建立EVPN邻居关系（比如：leaf与Spine，Spine与GW）来实现VxLAN隧道自动化建立，通常Spine作为RR来传递Overlay网络路由信息。
- DVS（vSwitch）通常作为纯二层交换机，实现安全组功能。
- DVS VM和SR-IOV VM均支持不同规格来承载VNFs的不同带宽流量。
- 开启ARP代答和主机ARP探测加速业务收敛，ARP和ICMP功能卸载到转发面，降低控制器CPU以及控制器和交换机之间管理通道带宽开销。
- 转发表项在交换机设备上保存，控制器离线后不会影响已建立服务通信。

4.2.3 软件 Overlay 组网

在软件Overlay组网方案中，OpenFlow作为Overlay网络集中式控制面，在VTEP和VxLAN GW之间建立VxLAN隧道。无需对Underlay网络改造，构建Underlay基础上的Overlay网络，支持多租户、大二层以及按需配置和自动化部署。



对于vSwitch做VTEP和vRouter GW，控制平面基于OVSDB和OpenFlow，通过OpenFlow TTP流表实现转发表项控制和主机位置学习。

DVS充当VTEP，V6000充当GW和NAT，支持多租户，V6000 GW采用主备方式部署，独占物理服务器，系统根据外部流量部署多对V6000 GW。每对V6000 GW按照外部网络的地址段进行调度，每个租户的出口只能落在一对V6000上，不同租户的网关可以落在不同的GW上，支持分布式路由，租户可以自定义安全组规则，通过控制器在DVS上下发流表实现。

4.3 跨数据中心互联

数据中心的网络可以划分为IDC内网和IDC外网。对于vDC网络方案来说，不仅需要构建高速互联、虚拟化、网络融合的IDC内网，同时需要提供完善的IDC外网服务。

对于IDC外网来说，有两个通道：服务发布通道和DC互联通道。

- 通过服务发布通道，可以安全、可靠、灵活接入。
- 通过DC互联通道，可以拓展网络边界、资源综合利用、迁移无感知、跨数据中心运算体验不下降。

vDC网络解决方案主要包括两类数据中心互联需求：

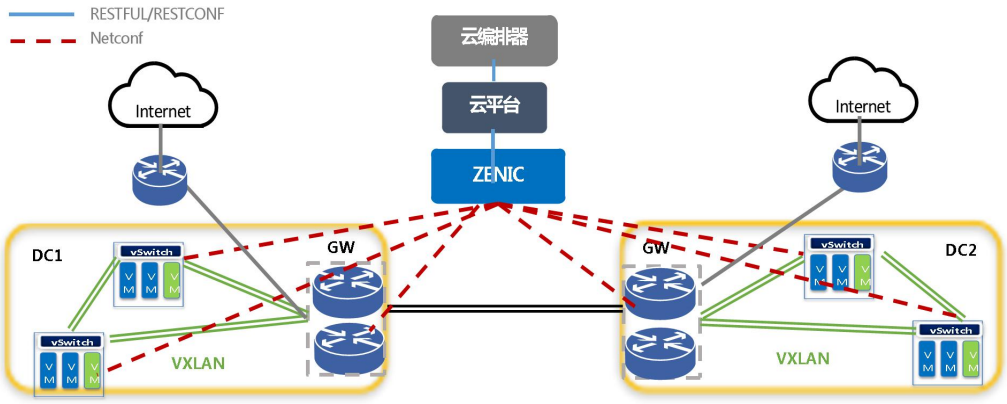
需求一：将多个小型IDC机房或分布在不同物理位置的零散机架资源整合到一个IDC，这个IDC即vDC，这样可以实现物理资源的多虚一整合。

需求二：vDC作为云数据中心的一种业务，在不同的IDC之间共享资源，实现云资源池的一虚多业务提供。

可以通过不同的数据中心互联技术，满足不同的应用场景。

4.3.1 单控制器多 DC 互联

一个编排器、一个云平台、一套控制器，管理多个物理DC，形成一个逻辑资源池。



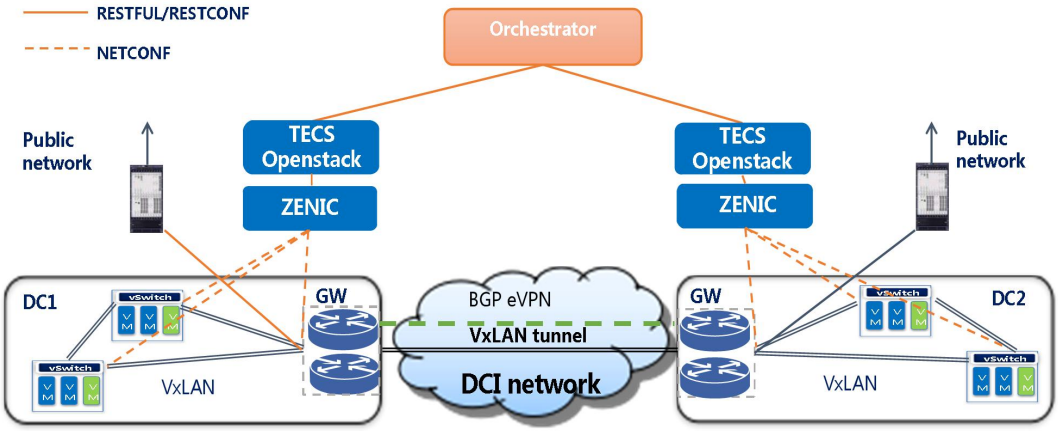
方案特点：

- 采用一套控制器集群系统，管理多个数据中心的转发设备，可以将整个多DC网络视为一个大的vDC。
- 数据中心内采用VxLAN网络构建Fabric虚拟化网络。

- 数据中心间采用基于VxLAN的互联方式，提供大二层的网络部署。
- 适用于多个小型IDC机房或分布在不同物理位置的零散机架资源整合到一个IDC，这样可以实现物理资源的多虚一整合。

4.3.2 多控制器多 DC 互联

一个编排器、多个云平台、多套控制器，网络通过EVPN技术构成一个逻辑资源池。



方案特点：

- 采用一套控制器集群系统管理一个数据中心的转发设备。
- 数据中心内采用VxLAN网络构建Fabric虚拟化网络。
- 数据中心间采用基于VxLAN的互联方式，提供跨数据中心的二层、三层网络部署。
- 数据中心间采用EVPN进行路由交互，实现MAC路由、IP路由的分发，避免跨数据中心互联中存在的广播、组播、未知单播流量。
- 适用于不同的IDC之间共享资源，或者跨SDN厂商的数据中心互通。

4.4 NFVI 解决方案

中兴通讯 SDN vDC 网络解决方案全面支持 NFVI，在基础的 SDN 网络基础上特别支持如下 NFVI 网络所需的特殊特性：

1. 对于高吞吐量 VNF 采用 SR-IOV 接入的支持，通过层次化网络方案支持每个 SR-IOV VF 分配唯一的 VLAN ID，在 ToR VTEP 上映射为指定的 VxLAN VNI。

2. 支持 VLAN Trunk，对于创建多个 VLAN 子接口的 VNF，可以通过 OpenStack+SDN 将指定 VNF 的 VLAN 映射到不同的 VNI 上。

3. VLAN Transparent，支持 VNF 自定义的 VLAN 互传，但是对于透传的 VLAN 中，不支持 MAC/IP 地址的重叠。

4. BGP As a Service 自动化，用于实现 VNF 和基础网络之间建立 BGP 邻居，并将 VNF 的路由发布给基础网络。中兴通讯的 SDN NFVI 方案采用集中式 BGP 控制面、分布式转发的方案，在 VNF VM 发生迁移后可以保证 BGP 邻居不断链。

5. BFD 自动化，实现 MANO 自动化编排 VNF 和基础网络之间的 BFD 配置。

5 vDC 网络关键技术

5.1 网络虚拟化

vDC网络技术，就是在已有的物理网络上，虚拟出来客户化的网络组件，因而实现网络的虚拟化。因此网络虚拟化是vDC网络技术的基础要求。

根据云数据中心的应用需求，网络虚拟化包括如下网络对象的虚拟化：

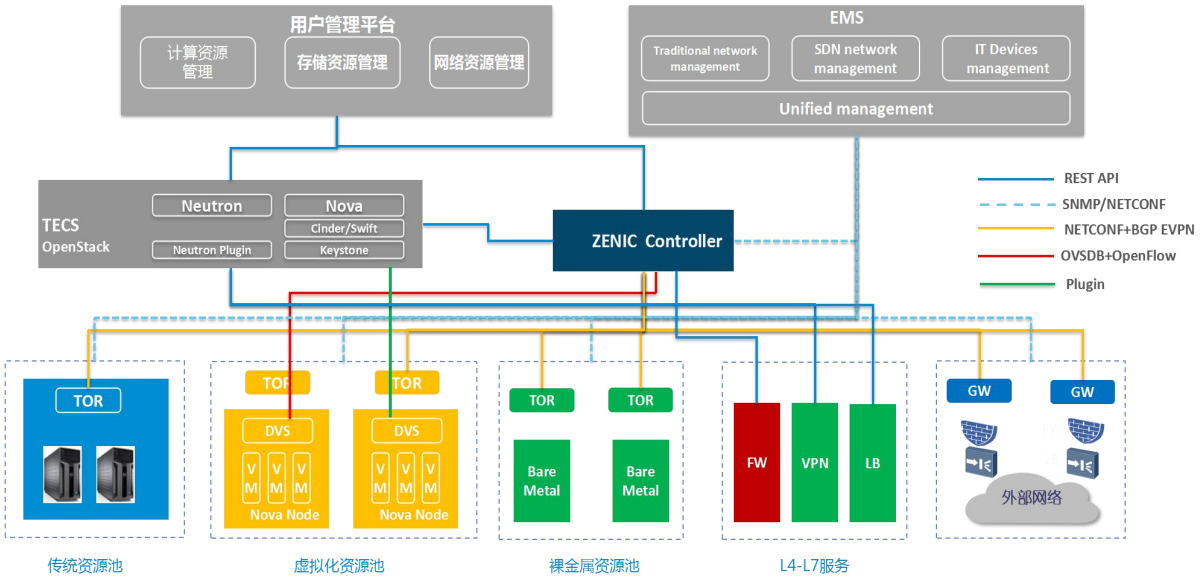
对象	虚拟化要求	虚拟化实现
Port	虚拟机端口虚拟化	vSwitch 或 SR-IOV 上的逻辑端口
网络	二层网络虚拟化	利用 VxLAN VNI 进行二层网络标识，实现二层隔离
子网	子网虚拟化	子网等价于路由前缀和 DHCP 地址池范围

路由器	路由器虚拟化	通过分布式路由功能提供东西向三层流量转发，当需要 NAT/Floating IP 服务时，通过集中式的 FW/NAT 设备提供服务
租户	租户虚拟化	通过 VRF 隔离，实现多租户的路由隔离，保证不同租户的访问隔离和地址空间可重叠
FW	防火墙业务虚拟化	对于采用 VNF 方式的 FW，可以为每个 FW 服务提供一个虚拟机，通过虚拟化软件实现 FW 的虚拟化 对于采用 PNF 方式的 FW，可以为每个 FW 服务提供一个 VSYS 系统，通过硬件设备的多实例实现 FW 的虚拟化
VPN	VPN 业务虚拟化	对于采用 VNF 方式的 VPN，可以为每个 VPN 服务提供一个虚拟机，通过虚拟化软件实现 VPN 的虚拟化 对于采用 PNF 方式的 VPN，可以为每个 VPN 服务提供一个 VRF，通过硬件设备的多实例实现 VPN 的虚拟化
LB	LB 业务虚拟化	对于采用 VNF 方式的 LB，可以为每个 LB 服务提供一个虚拟机，通过虚拟化软件实现 LB 的虚拟化 对于采用 PNF 方式的 LB，可以为每个 LB 服务提供一个 VRF，通过硬件设备的多实例实现 LB 的虚拟化

vDC网络技术包括基础的L2/L3网络、子网、安全防护、VPN、负载均衡器等等，充分满足云数据中心的各类网络需求。

5.2 系统接口和协议

中兴通讯SDN数据中心网络解决方案实现网络虚拟化，ZENIC vDC控制器是该解决方案中最重要的组件，它支持丰富的接口和协议。



- 南向接口方面，ZENIC控制器支持OpenFlow1.3/1.5，BGP eVPN，OVSDB，BGP，NetConf/YANG，SNMP等协议，南向可以纳管的设备包括硬件和虚拟化网络设备。如：OVS，DVS，OpenFlow交换机和路由器，传统交换机和路由器。
- 北向接口方面，ZENIC控制器可以为SDN应用提供REST/RESTconf接口，SDN应用包括TECS/Openstack、EMS、VNFM等。ZENIC控制器提供Neutron插件与Openstack集成。
- 对于传统路由器、交换机，ZENIC控制器通过标准的路由协议对路由器、交换机进行配置，所使用协议包括NetConf/BGP eVPN等。
- 对物理防火墙，ZENIC控制器通过REST协议管理。

5.3 L4-L7 网络功能虚拟化

vDC网络虚拟化，根据实现方式的不同，可以分为两类：

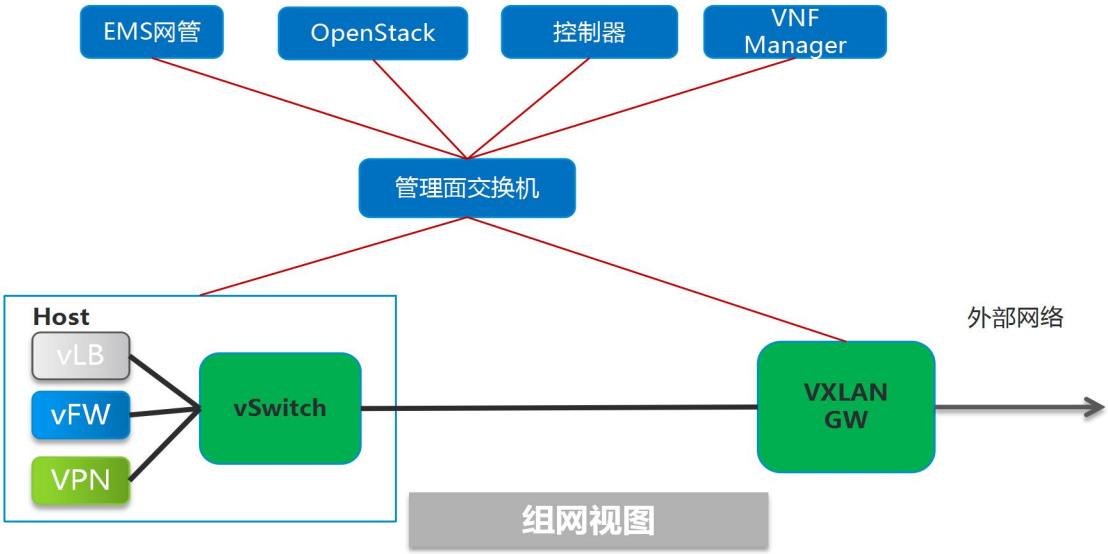
- 基础网络虚拟化，主要是指端口虚拟化，网络虚拟化，子网虚拟化，以及路由虚拟化等基础网络服务的虚拟化。
- L4-L7网络功能虚拟化，主要是更高层次的网络服务虚拟化，比如FWaaS，VPNaaS，LBaaS等L4-L7网络服务。

L4-L7网络功能的虚拟化可以分为基于VNF（Virtual Network Function）的虚拟化实现方式和基于PNF（Physical Network Function）的虚拟化实现方式。

1. VNF虚拟化

基于VNF（Virtual Network Function）的虚拟化实现方式，是指以虚拟机的方式部署L4-L7网络服务，比如将FW软件部署在虚拟机中，每一个虚拟机就是一个虚拟防火墙，为租户提供防火墙服务。

在以VNF虚拟化方式提供L4-L7服务时，需要包括如下组件：



1）VNF Manager

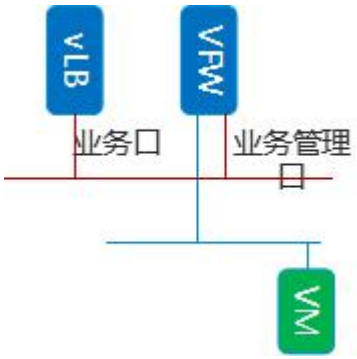
VNF Manager即VNF管理网元，主要提供两方面的服务。

- VNF的生命周期管理，包括创建VNF虚拟机、删除VNF虚拟机、修改VNF虚拟机等等，会调用OpenStack的接口进行虚机动态创建/删除。
- VNF的策略配置，主要负责将在云管理平台设置的L4-L7服务策略（比如防火墙的安全策略，过滤规则等）转换成VNF的配置，并下发到VNF中。

2）VNF

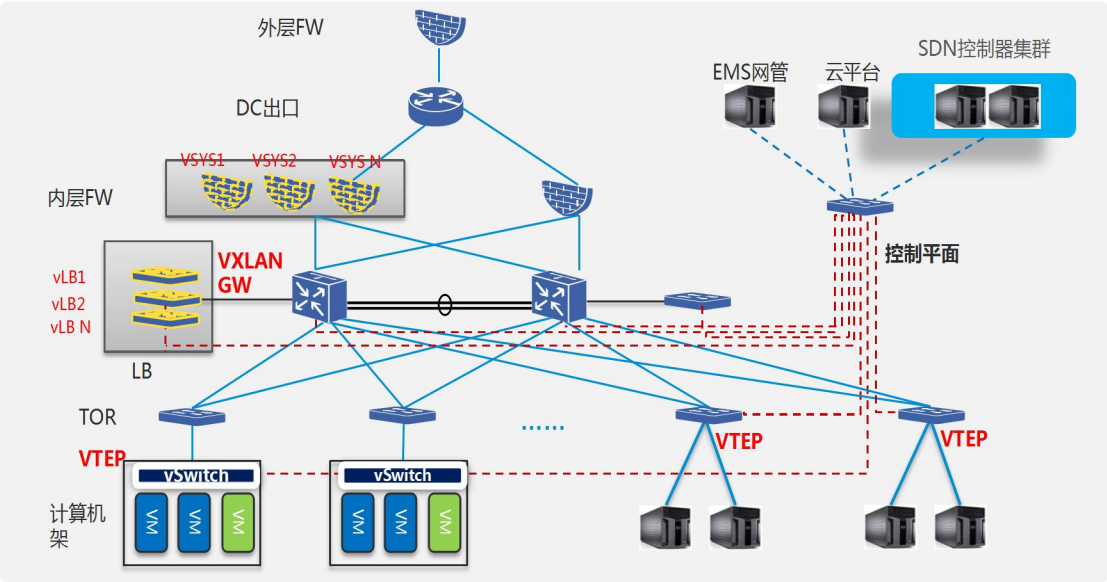
VNF即提供L4-L7网络功能的实体，从功能上可能包括了LB、FW、VPN等服务。VNF一般提供两个网络接口：业务口和管理口。

- 业务口主要负责流量转发和处理。
- 管理口主要与VNF Manager进行通信，VNF Manager通过管理口对VNF进行管理。



2. PNF虚拟化

基于PNF（Physical Network Function）的虚拟化实现方式，是指以物理设备部署L4-L7网络服务，并通过设备的一虚多服务提供网络功能的多实例，比如将硬件FW设备按照VSYS进行划分，每一个VSYS就是一个虚拟防火墙，为租户提供防火墙服务。



硬件设备一虚多的实现方式根据不同的产品有所差异，常见的包括VSYS、VRF等方式。

1）VSYS

VSYS是一种物理硬件一虚多的技术，它能够将一台物理设备在逻辑上划分成多个虚拟设备，每个虚拟设备系统都可以被看成是一台完全独立的设备，拥有独立的系统资源，且能够实现设备本身的大部分功能。

每个虚拟设备系统之间相互独立，不可直接相互通信。

- 每个VSYS拥有独立的管理员
  每个VSYS拥有独立的安全域、地址簿、服务簿等
  每个VSYS可以拥有独立的物理接口或者逻辑接口
  每个VSYS拥有独立的安全策略
- 每个VSYS拥有独立的日志

2）VRF

VRF(Virtual Routing Forwarding)是一种路由虚拟化技术，设备上的每一个VRF可以看作虚拟的路由器，好像是一台专用的设备。每个VRF包括如下元素：

- 一张独立的路由表，当然也包括了独立的地址空间
- 一组归属于这个VRF的接口的集合
- 一组只用于本VRF的路由协议

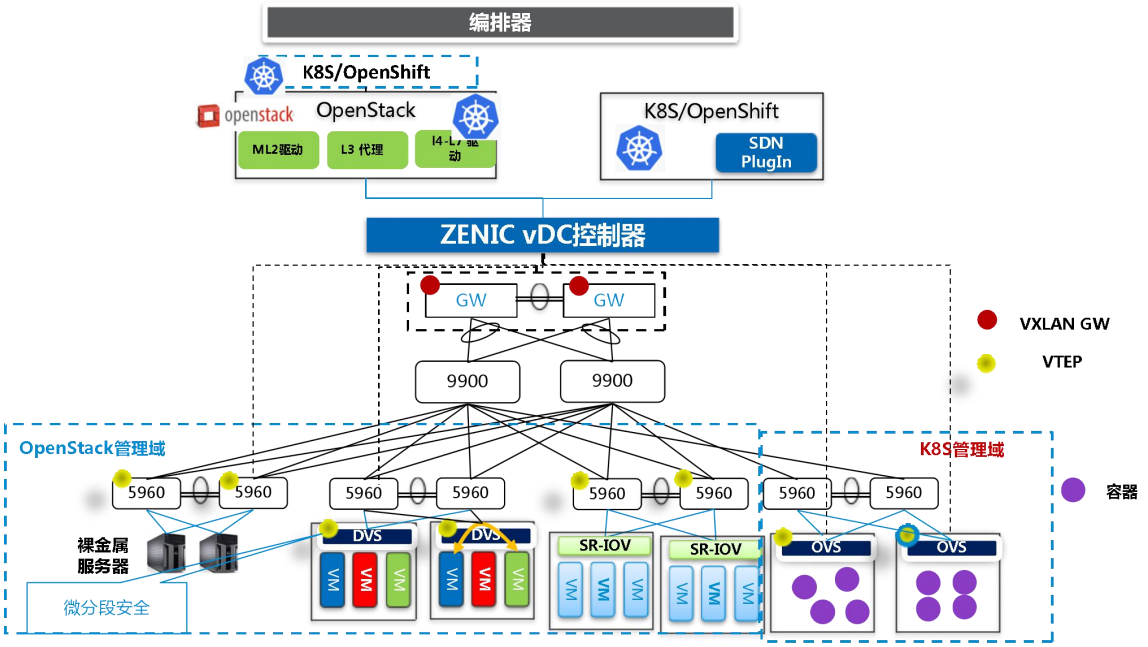
使用VRF可以实现设备路由的一虚多，从而实现多VRF的地址重叠问题。

物理L4-L7设备可以采用VRF提供设备虚拟化服务。

5.4 多类型计算负载统一管理

中兴通讯 vDC 解决方案统一管理虚拟化服务器、裸金属物理服务器、SR-IOV 虚机、容器等多种计算负载，并且可以获得统一自动化业务发放、集中管控的特性。对于虚拟化、容器服务器。采用 DVS 虚拟交换机接入

虚拟化服务器，对于 SR-IOV、裸金属服务器，采用硬件交换机 5960 VTEP 统一接入。支持 OpenStack 和 K8S 共享一套网络。支持一个租户同时申请虚机、容器、裸金属，网络资源分配不受计算资源形态的影响，其网络资源的关联由编排器编排，SDN 控制器完成关联。

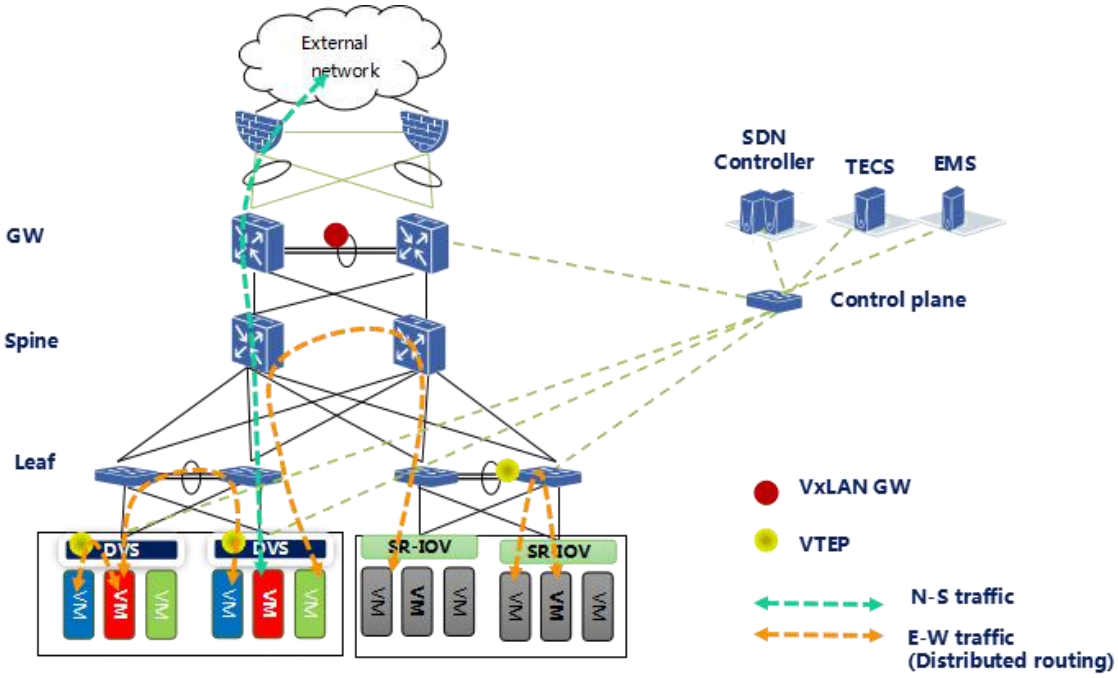


5.5 全分布式路由方案

分布式路由实现VM或裸金属服务器互访流量从最近接入点进行三层转发，无论是DVS VTEP还是硬件交换机5960作为VTEP，均可以全自动化实现分布式路由，流量无需经过集中的网关点。这样带来优势有：

- 流量路径的最优化，无需经过网关迂回，同一租户内的跨子网流量路径减少数跳，大大减少了流量的时延。
- 流量路径的优化也使得同等的业务流量模型下，GW的投资减少一半以上，并且对于核心交换机到GW、ToR到核心交换机的带宽要求也大大下降，避免了业务高峰时，网络成为瓶颈。

如图所示，流量转发路径如下：

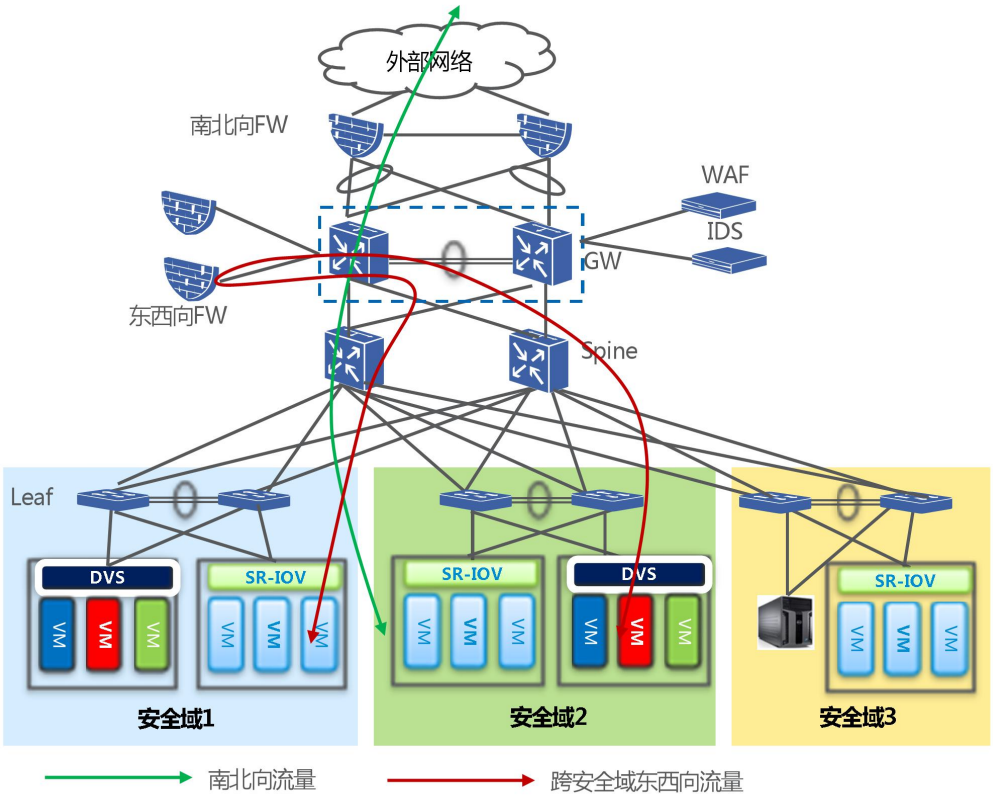


- 如果两通信主机位于同一VTEP下，流量将通过VTEP路由，不会将流量转发至GW。
- 如果两通信主机位于同一Leaf不同vSwitch下，流量将通过Leaf转发，不会将流量转发至GW。
- 如果两通信主机位于不同Leaf下，流量将通过Spine转发，不会将流量转发至GW。
- 只有南北向流量会通过GW进行转发。

5.6 网络安全方案

- 多租户隔离安全：自动的多租户安全隔离能力，不同的租户之间的流量、地址空间完全隔离。
- 边界安全防护：部署南北向防火墙，提供租户级别的边界安全防护，将外部的安全攻击阻断在边界上。
- 同租户安全域隔离：支持自动化安全域划分，当入侵者攻破非信任域主机时不能威胁核心系统数据的安全。

- 东西向安全隔离： 提供了微分段安全，支持在同租户内进行安全边界的划分，实现有状态的安全组。
- 优先级带宽限速：每个租户配置出口优先级和带宽限速，确保不会阻塞更高优先级业务的出口。



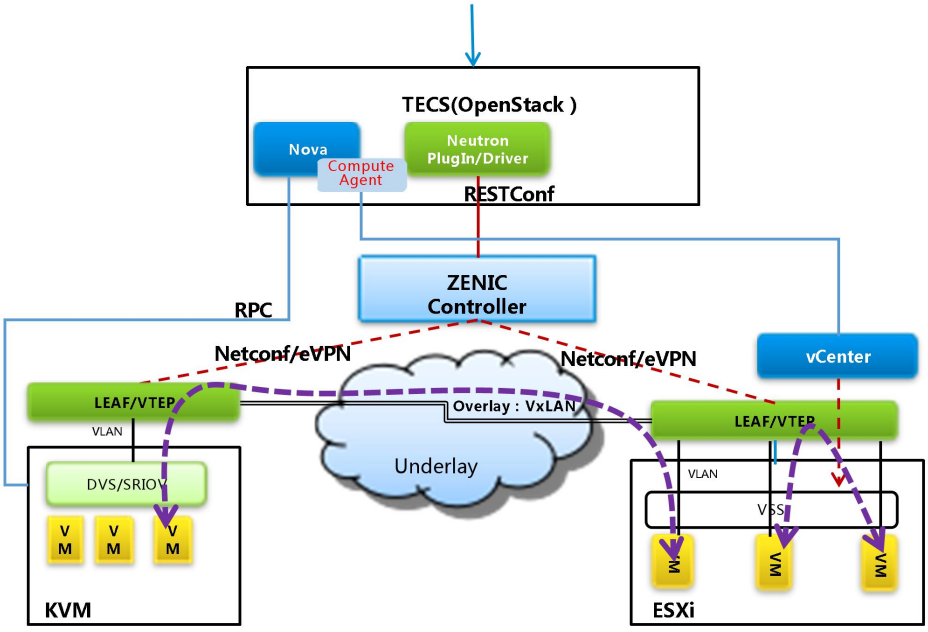
5.7 KVM 与 VMWare 统一纳管方案

控制器支持层次化端口方案，ZENIC vDC 控制器和云平台 OpenStack 的对接方案如下：

VMWare 兼容方案通过提供 Neutron 插件 VMWare Driver 和 Nova 插件 Compute agent 与 OpenStack 进行集成。整体网络方案采用层次化端口方案，SDN Controller 负责 VxLAN 相关业务配置，云管系统负责底层 VLAN 网络配置，方案中主要涉及 Neutron 插件包括了 ML2，L3 Agent，LB Driver，FW Driver 等。

Neutron SDN Plugin 以 ML2 Driver/L3 Agent 的形式存在，运行在 Neutron 控制节点上，将 Neutron 基础网络接口调用转换成 SDN 的北向接口消息发送给 ZENIC 控制器，由后者转换为对网络的各类控制消息。基于 ML2 Plugin 的层次化端口方案，SDN Controller 负责 VxLAN 相关业务配置，并采用层次化端口绑定技术实

现 VLAN-VxLAN 的映射。



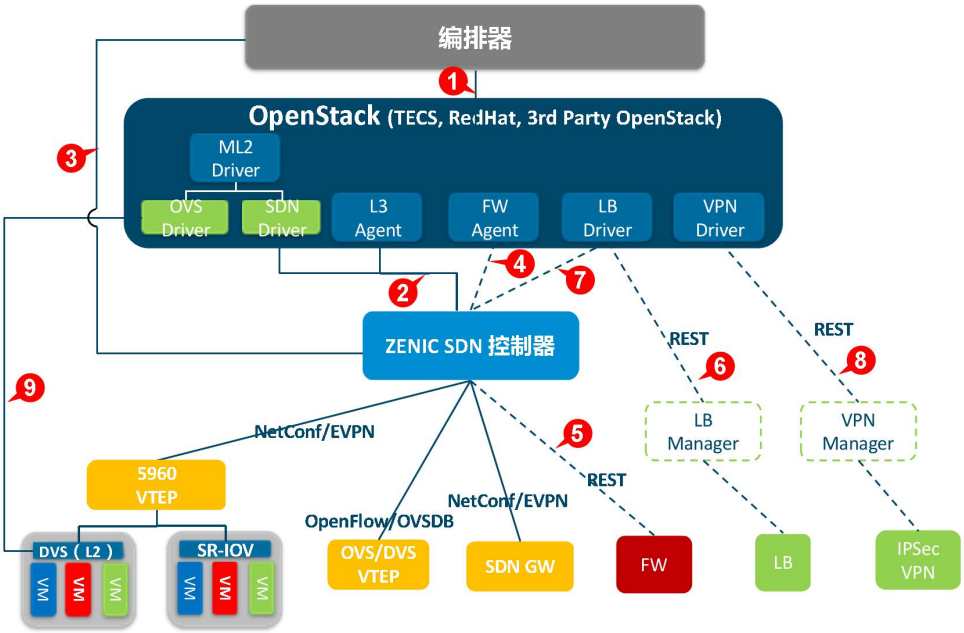
网络资源发放过程 - 创建 VM 过程

- 控制器上导入ESXI主机和TOR端口的连接关系。
- 租户通过Openstack创建Network，并通知到SDN Controller。
- 租户通过Nova上Compute agent调用vCenter在此Network下创建VM。
- VSS感知VM上线，上报Openstack，Neutron调用VMware Driver为此VM的端口申请VSS上的VLAN ID，并通过旁路使VSS VM直通Leaf交换机。
- ML2将VMware driver的VLAN分配信息通过SDN driver通知到SDN Controller。
- SDN Controller为VM所在ESXI主机连接的TOR端口上下发相关网络配置，建立VLAN和VxLAN间的映射关系。

通过层次化端口方案，可以实现 EXSi 主机上的 VM 的 VLAN 和 VxLAN 的映射关系，从而实现 ESXi 主机和 KVM 主机之间统一的 VxLAN 网络互通。

5.8 Openstack 云网联动方案

中兴通讯 vDC 网络解决方案可用于电信云、私有云、公有云、混合云等多种业务场景。整体解决方案包括 TECS/Openstack 统一云管理平台、ZENIC SDN 控制器、Spine-Leaf 交换机、SDN 网关和其他组件（L4-L7 层服务：FW、LB 等），可提供混合云组网、多 DC 组网，实现异构资源池纳管和数据中心智能运维。



接口功能：

① 编排器和 OpenStack 接口基于标准 OpenStack 北向 API，Openstack networking API。

② ZENIC SDN 控制器北向接口，由控制器提供 Neutron Plugin 插件（包括：ML2 drier、L3 agent、FW driver、LB driver、VPN driver 等）与 OpenStack 的 Neutron 对接。

③ 部分订制功能，编排器直接调用控制器接口。

④ FW driver 调用控制器接口打通基础网络到 FW 的流量路径，硬件防火墙要创建 VLAN 子接口对应 FWaaS 实例。

⑤ FW driver 实现对 FWaaS 业务的开通、策略编辑。

⑥ LBaaS 业务开通，由 LB Driver 自己实现。

⑦ LBaaS 实例和基础网络之间的配置，VNF 免配置，PNF 要配置 VLAN 子接口和 LBaaS 实例之间的

对应管理。

⑧ VPNaaS 实例配置，VNF 模式下无需和网络交互，指向 VPN 的路由也通过调用标准 TECS/OpenStack 接口实现。

⑨ RPC：云平台和 vSwitch 以及 SR-IOV agent 接口。

SDN 提供 OpenStack 插件,将 Neutron 调用转换为控制器、LB/FW 的 REST 调用。

VTEP：OVS/DVS 使用 OVSDB 作为管理面,采用 OpenFlow 控制转发表项；对于硬件交换机采用 NetConf 作为管理面。

L4-L7 Service：控制器采用 REST 接口纳管第三方防火墙，其它提供插件直接配置设备。

5.9 层次化端口方案

对于VTEP在TOR交换机的场景（裸金属服务器、SR-IOV服务器或者OVS只做L2交换机），中兴采用层次化端口方案，层次化端口指ToR交换机上行采用VxLAN组网，服务器到ToR之间采用VLAN组网，采用VLAN唯一标识虚机。SDN Driver负责将所有网络请求发往SDN控制器，OVS Driver会读取SDN插件分配的VLAN，并将其写入到OVS的配置中（如果是SR-IOV场景，则由SR-IOV的插件负责将VLAN信息写入到网卡中）。

5.10 资源映射关系

TECS/OpenStack的网络模型和SDN网络的映射关系如下表所示：

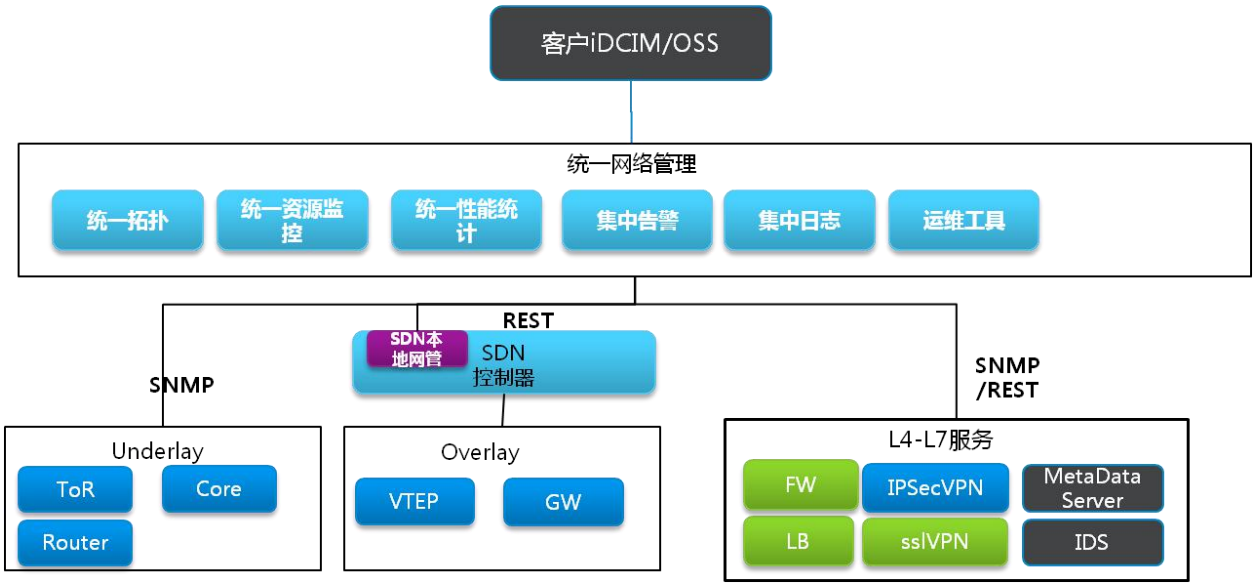
TECS/OpenStack 对象	SDN 网络对象
Tenant/project	无，控制器上仅表示为一个顶层容器对象
Network	L2 VNI 标识二层网络标识，实现二层隔离， VNI 值=segment id
Router	VRF, 通过 L3 VNI 标识，通过分布式路由功能提供三层流量转发
Router Interface	虚拟 L3 接口，控制器在制定 VRF 下创建 IRB 接口作为 VM 的三层网关，并绑定 L2 VNI
port	Port-VNI 的映射关系，对应于 vSwitch 的逻辑端口或 TOR VTEP 的物理端口

Subnet	DHCP 地址池和相关 DHCP 选项
Security groups	ACL 规则，有状态规则以实现微分段隔离
FWaaS	逻辑防火墙实例或者 vFW
LBaaS	逻辑负载均衡器实例或者 vLB
VPNaaS	采用 VNF 方式的 VPN，可以为每个 VPN 服务提供一个虚拟机

5.11 统一网络管理

云数据中心采用了Overlay技术承载网络的虚拟化，整个IDC网络中不仅包括了传统的Underlay设备，比如交换机，路由器，防火墙等设备，还包括了Overlay层次的虚拟交换机，VxLAN GW，TOR VTEP交换机，VNF/PNF形式的L4-L7设备，给网络管理带来的新的挑战。

vDC网络解决方案采用统一网络管理方案，实现多层次的网络设备的管理和监控。



- 拓扑呈现

统一网络管理可以支持虚拟拓扑、物理拓扑的统一拓扑视图。

- 日志管理

统一网络管理可以支持日志管理功能，用户操作，网络事件，系统日志进行记录。

- 告警管理

统一网络管理可以支持各层面的集中告警管理。

- 运维工具

统一网络管理可以支持端到端的故障诊断功能。

- 性能统计

统一网络管理可以支持性能统计，能够对虚拟网络、子网、虚拟路由器等虚拟网络对象，以及Underlay层次的控制器、NVE、网关、虚拟防火墙、虚拟负载均衡器等网络对象的关键指标进行性能统计。

- 资源监控

统一网络管理可以支持资源监控，可以监控虚拟资源的使用情况，以及硬件设备/物理资源的使用状态。

6 vDC 网络技术总结

vDC网络技术是指在计算、网络硬件基础上，通过构建虚拟化数据中心网络，为数据中心中承载的云应用提供网络服务，包括基础的L2/L3网络、子网、安全防护、VPN、负载均衡器等等

vDC网络技术，可以提供完善的云数据中心解决方案：

- 网随云动
    - 逻辑路由器、L2网络、QoS、安全策略、FWaaS、LBaaS即刻下发到网络，即时生效，避免任何人工配置
    - 虚机迁移，策略跟随
- 多租户、自服务
    - 网络分区、隔离，支持多租户网络

- Network As a Service，用户自服务，网络维护和业务维护解耦

- 可视化，简化运维
  - 拓扑可视化
  - 流量可视化
  - 故障可视化
- 软件定义数据中心网络
  - 开放接口，用户可编程
  - 新功能需求、运维需求，用户可开发，加快市场响应速度

通过vDC网络技术，构建公有云、私有云、混合云、NFV基础设施云等各种新型的数据中心，满足多样的网络需求。