数据中心虚拟机网络接入技术——应用篇

当前企业数据中心由于面临着服务器数量与日俱增、数据量日益庞大、管理维护困难等诸多挑战，网络扁平化、服务器虚拟化、数据中心 IP 网络与存储网络融合等技术成为关注的重点。虚拟化技术可以解决业务快速部署与灵活迁移、资源共享的问题，同时也带来了虚拟机与虚拟交换网络的部署与管理维护的问题。IRF、VLAN 等网络虚拟化技术作为虚拟化管理的另一方面，已经相对比较成熟，相比服务器虚拟化不会给网管系统带来新的冲击。本文主要关注服务器虚拟化技术给企业数据中心管理带来的问题：

- **职责不清：**虚拟交换机（vSwitch）的出现，使网络触角已经延伸到服务器中，系统管理员和网络管理员的职责分界不清(注：关于 vSwitch 及虚拟机相关网络概念的介绍请参照本刊《数据中心虚拟机网络接入技术——基础篇》一文)；
- **资源可视性：**服务器在虚拟化后，规模剧增；虚拟机网络位置的非物理性，使资源的可视性管理非常困难。虚拟机的存在给故障（网络、应用）分析和定位带来更大的难度；
- **流量可视性：**虚拟机的迁移技术，使网络中的流量分布存在更多的动态性，导致流量可视性的困难；
- **自动化配置管理：**虚拟机的迁移技术，要求网络的配置要跟随虚拟机进行动态实时的调整、自动化管理能力的必要性更为突出；

下面就针对数据中心管理中的这些问题进行分析同时提供一定的解决方案。

一、物理和虚拟网络一体化的资源可视性管理

服务器虚拟化后，虚拟服务器规模剧增，以及虚拟化软件的迁移特性使虚拟服务器在数据中心网络中的物理位置的可视性变得困难。

当业务系统异常时，需要从服务器、网络各方面进行分析诊断，对网络管理员来讲，需要清楚虚拟服务器 VM 位于哪个物理服务器、通过哪个物理网络交换机接口接入网络，甚至需要了解 vSwitch 上的网络配置（比如 VLAN），特别是服务器和网络的边界链接的可视性。如果对这些信息无法可视化管理，就无法有效的分析和定位故障。

当前的解决方案是借助虚拟化软件（如 VMWare ESX/ESXi）对服务器、VM 的管理能力，将网络相关信息纳入到统一的资源可视化管理中。

- **虚拟资源视图**

如图 1 所示，虚拟资源视图能力为我们提供了物理服务器、虚拟交换机、VM 的资源从属关系信息。

图1 H3C iMC虚拟资源视图

如图 2、图 3 所示，虚拟交换机的管理能力，提供了服务器中虚拟网络的配置能力（端口数量、端口组、VLAN、和物理网卡的绑定关系等）。

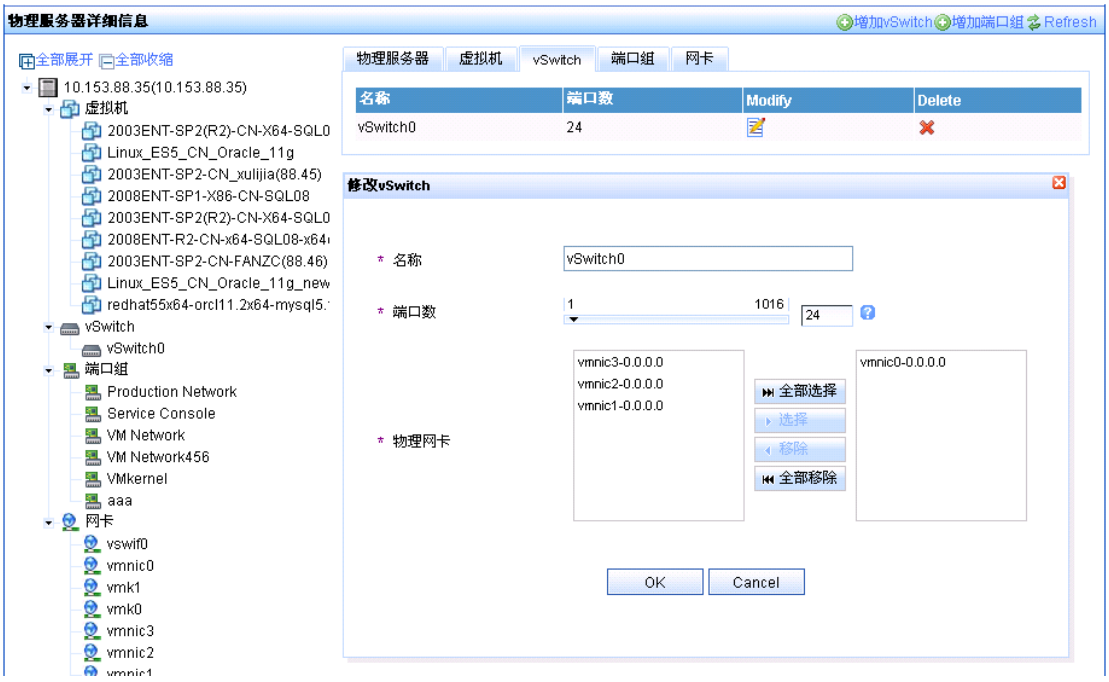

图2 H3C iMC虚拟交换机管理

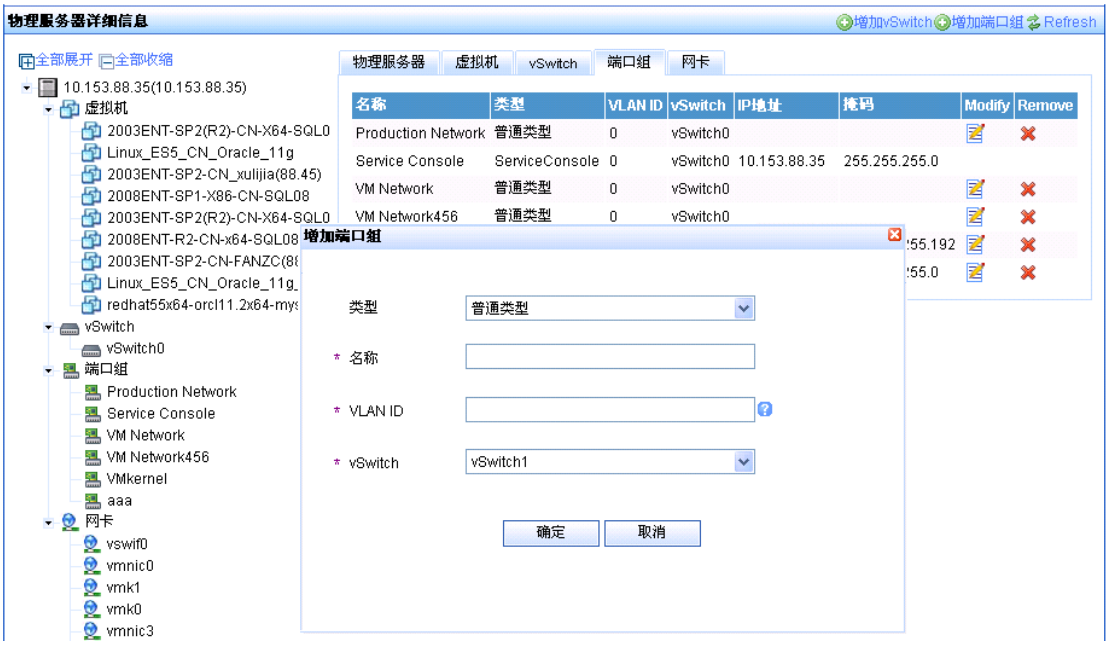

图3 H3C iMC虚拟交换机端口组管理

如图 4 所示，对虚拟服务器 VM，提供分配的计算资源、GuestOS 信息的可视性。

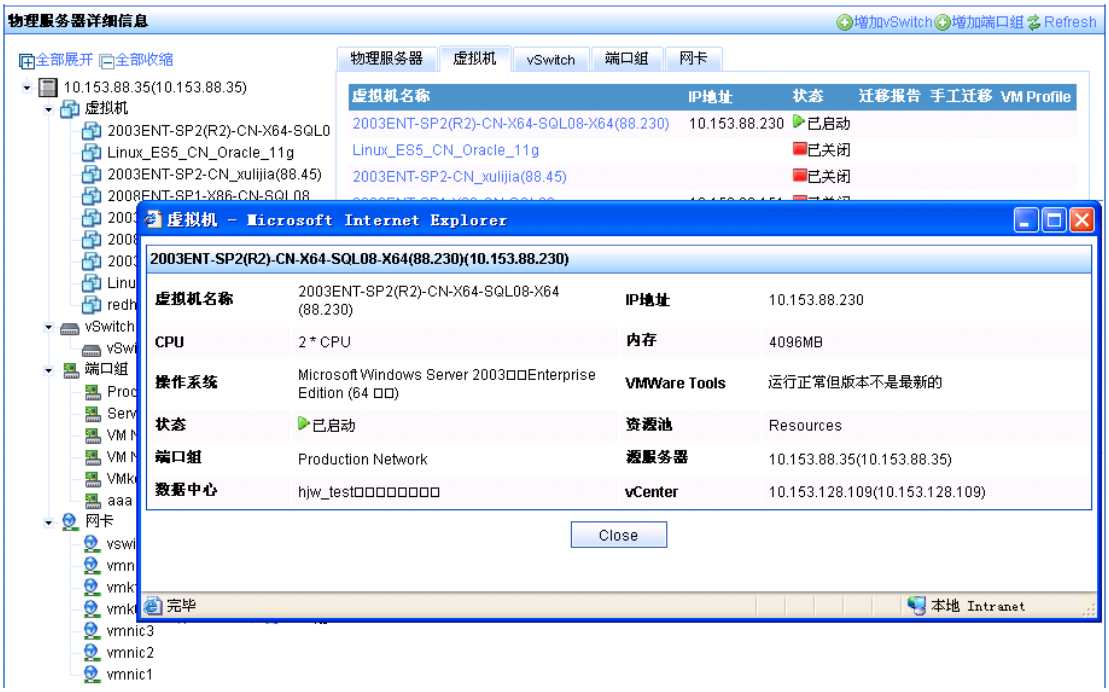

图4 H3C iMC虚拟机信息管理

- **虚拟网络拓扑**

拓扑是最为直观的管理方式，通常的网络拓扑由于没有计算虚拟化相关数据，无法意识到各个虚拟服务器 VM 和物理服务器、vSwitch 的从属和链接关系，各个 VM 是零落到整个拓扑、不同网段中的独立节点，而且 VM 的数量远远大于物理服务器的数量，最终出现的是一个大量、杂乱、无关的拓扑。

如图 5 所示，虚拟网络拓扑为解决这个问题，在拓扑计算中使用虚拟网络的拓扑数据，提供清晰简洁的物理拓扑，所有虚拟节点都聚合到物理服务器节点上，同时又能体现物理服务器内部的虚拟世界。

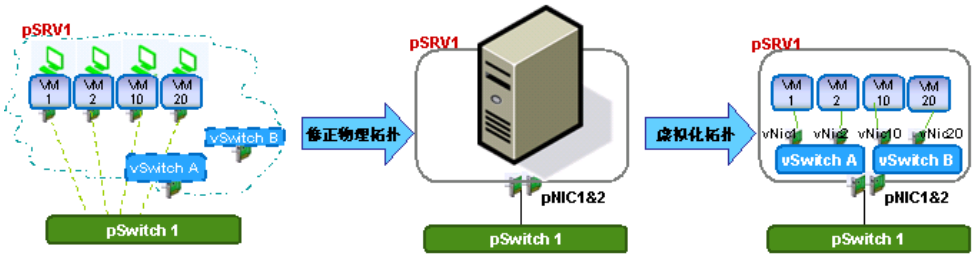

图5　虚拟网络拓扑的目的

虚拟网络拓扑的展现如图 6 所示，展示物理服务器（ESX）、虚拟交换机（vSwitch）、虚拟机（VM）之间的从属或连接关系；同时，通过 ESX 和物理交换机之间的连接关系，展示 ESX 所在的物理位置。

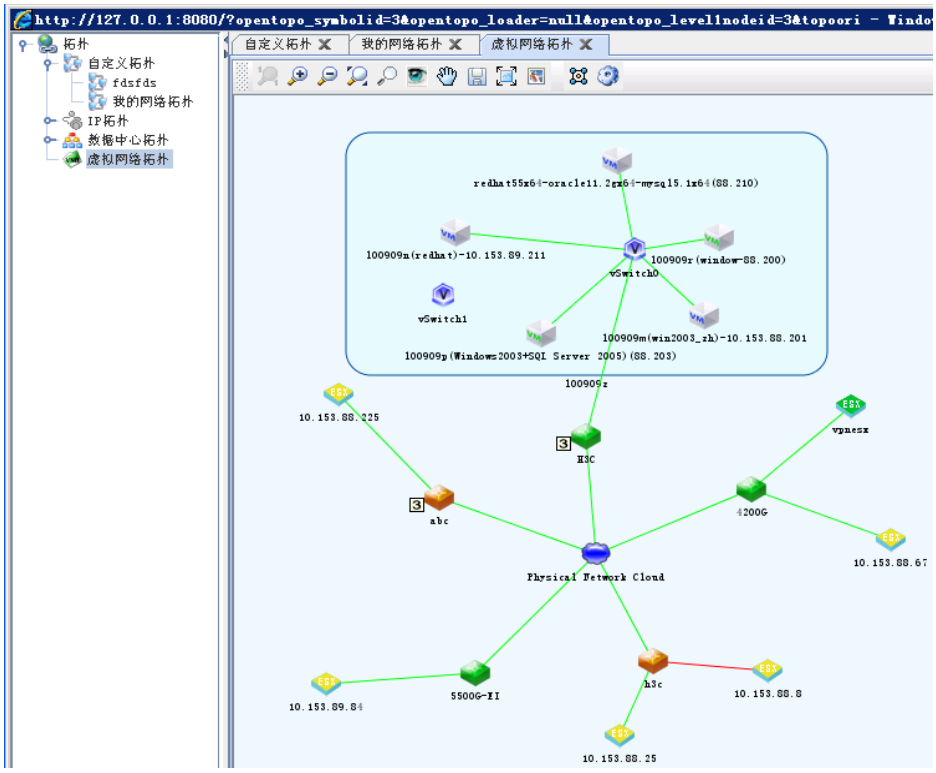

图6　虚拟网络拓扑实例

二、物理和虚拟网络一体化的流量可视性管理

企业数据中心不仅仅存在着外部对数据中心应用的访问流量，在数据中心内部应用之间反而存在着更为大量的数据交换，掌握这部分流量的分布以及对网络的需求，对保障其业务的正常运行有更大的意义。如图 7 所示，展示了数据中心内多个应用业务之间的流量。简单的来看，就是采用流量分析技术（NetStream/SFlow， 镜像流量/探针），以服务器地址组+四层端口号组作为数据中心内业务流量的标识，进行业务间流量分析。

A-B	B->A流量	B->A流速	A->B流量	A->B流速	关注
订单-订单	4.31 GB	10.29 Mbps	4.31 GB	10.29 Mbps	取消关注
订单-支付	0.61 GB	1.46 Mbps	0.83 GB	1.97 Mbps	取消关注
订单-搜索	0.33 GB	0.79 Mbps	0.40 GB	0.96 Mbps	取消关注
其他业务-订单	0.40 GB	0.94 Mbps	0.24 GB	0.57 Mbps	取消关注
支付-支付	0.12 GB	0.28 Mbps	0.12 GB	0.28 Mbps	取消关注
支付-搜索	63.43 MB	0.15 Mbps	55.08 MB	0.13 Mbps	加入关注
其他业务-支付	66.98 MB	0.16 Mbps	45.44 MB	0.11 Mbps	加入关注
其他业务-搜索	27.73 MB	64.62 Kbps	34.48 MB	80.34 Kbps	加入关注
搜索-搜索	27.74 MB	64.63 Kbps	27.74 MB	64.63 Kbps	加入关注

图7　应用流量可视性举例

如果应用是部署在虚拟机 VM 上的，就需要考虑虚拟机 VM 之间的流量如何感知和获取。传统虚拟化技术中的 VEB vSwitch 模式，无法支持流量的可视化管理能力。虚拟机 VM 之间的相互流量直接在 vSwitch 上交换，网络是无法感知的（如图 8 所示），而传统的 VEB vSwitch 本身缺少内部流量监管能力，比如端口报文统计、端口流镜像、NetStream 等。因此通过传统的网流分析手段，无法完成流量可视化的目的。

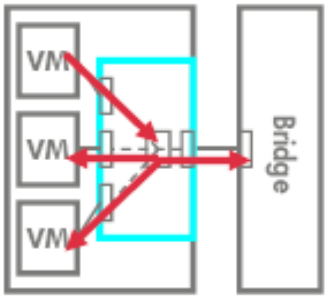

图8 VEB模式流量模型

目前正在形成标准的 VEPA 模式（以及 Multi-Channel、Port-Extender），VM 间的流量必须通过外部网桥进行交换（如图 9 所示），网络具有完全的流量可视性，只要网流分析管理软件能够将触角延伸到 VEPA 外部网桥上即可。该模式要求对应网桥支持 NetStream、NetFlow、SFlow 能力。(注：关于 VEPA 概念的介绍请参照本刊《EVB——数据中心虚拟机网络接入技术》一文)

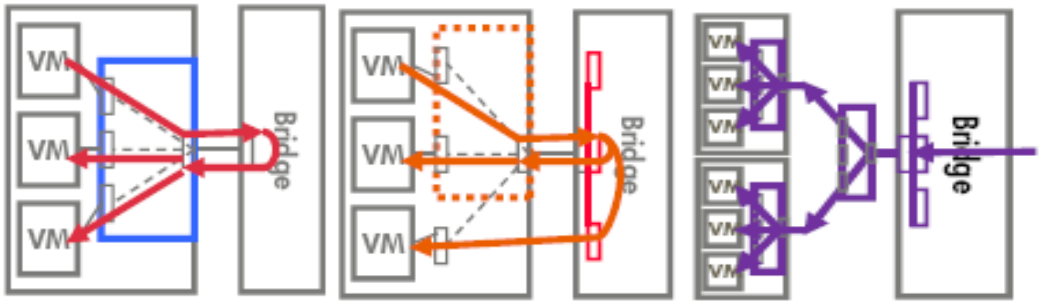

图9 VEPA、Multi-Channel、PE模式流量模型

三、随需而动的自动化配置管理

在创建 VM 或迁移（vMotion）时，VM 主机能否正常运行，不仅需要在服务器上的资源合理调度，网络连接的合理调度也是必须的。图 10 所示虚拟机 VM1 从物理服务器 pSrv1 迁移到物理服务器 pSrv2 上。

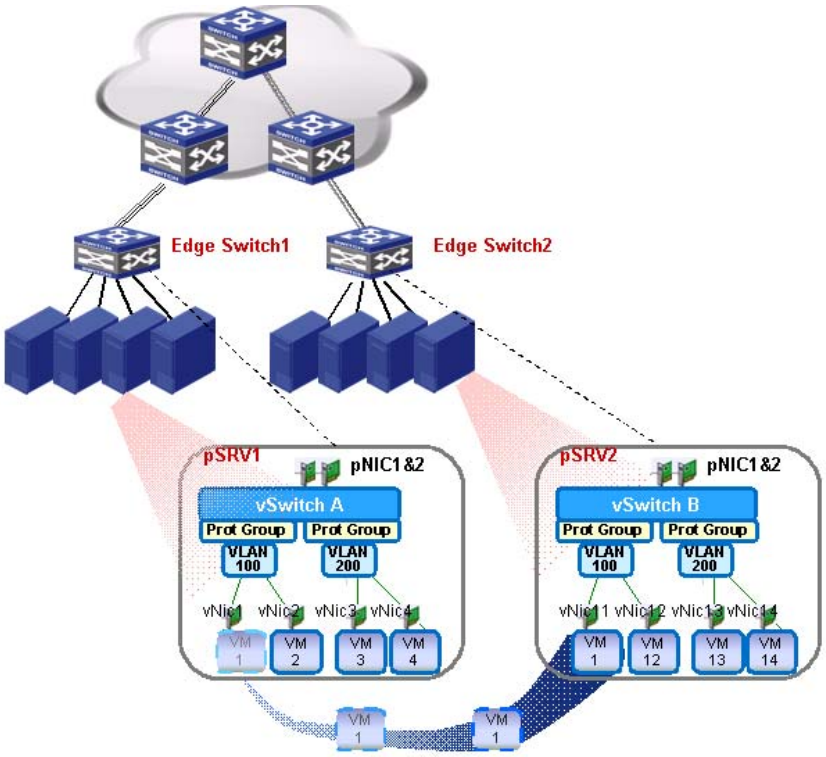

图10 网络配置迁移

其网络连接从原来的由 pSRV1 上虚拟交换机 vSwitchA 的某个 VSI(属于 VLAN100 的端口组)接入到边缘物理交换机 Edge Switch1，变成由 pSRV2 上 vSwitchB 的某个 VSI(属于 VLAN100 的端口组)接入到 Edge Switch2。若迁移后对应的 Edge Switch 的网络配置不合适，则 VM1 迁移后就可能不能正常使用。比如原先对 VM1 的访问设置了 ACL，以屏蔽非法访问；或设置了 QoS，以保障 VM1 上业务运行带宽等服务质量。都需要在发生 VM 创建或 vMotion 时同步调整相关的网络连接配置。而且，为了保证 VM 的业务连续性，除了虚拟化软件能保证 VM 在服务器上的快速迁移，相应的网络连接配置迁移也需要实时完成。即网络具有“随需而动”的自动化能力。

1. 基于 VMware vCenter 的配置迁移方案

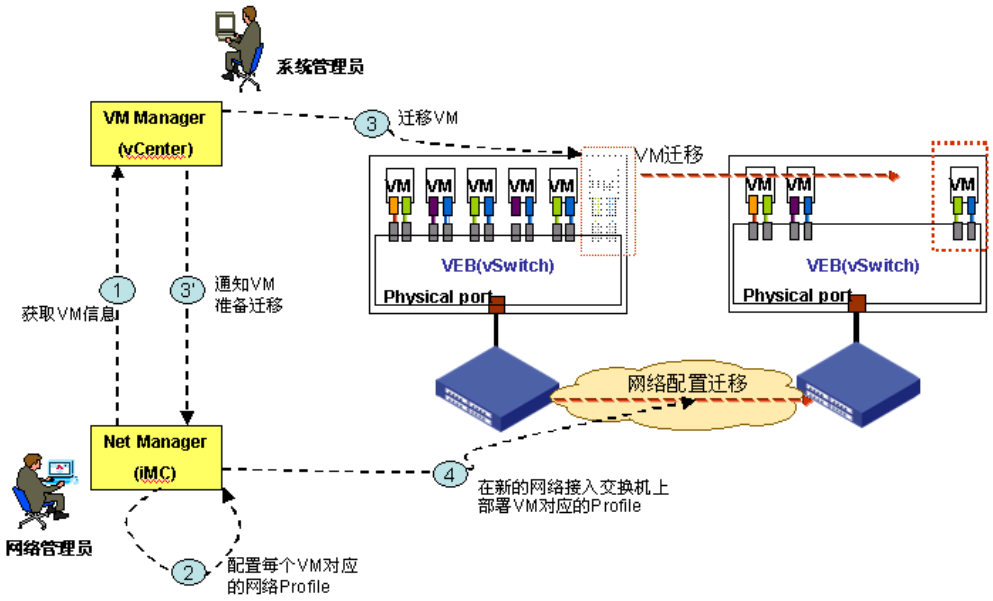

图11 基于VMware的虚拟网络配置迁移原理

① 网络管理员通过虚拟资源可视化能力，获取虚拟机的清单信息（包含每个 VM 使用的 vNic，可能有 1 个或多个）

②网络管理员针对每个 VM 定义对应的网络配置 Profile

③系统管理员在 vCenter 上触发 VM 迁移，并通知网络管理系统

④网络管理系统将对应的网络配置 Profile 下发到本次迁移新的网络接入交换机上。这里也要借助网管系统对虚拟资源的可视化管理能力，定位出 VM 连接到的物理网络交换机的接入位置。

2. Multi-Channel、PE 提供了更优的配置迁移方案

在 VEB vSwtich 模式下，多个 VM 的 vNic（对应 VEB 上的 VSI 接口）和邻接物理交换机的链接使用的是一个物理链路（或聚合后的物理链路，Nic teaming），不管是逻辑上还是物理上都是一个接口，即 VM 和物理接口是 N:1 的关系（如图 12 所示）。

在此模式下，邻接物理交换机的配置控制粒度只能到物理接口级，针对数据中心“随需而动”的配置自动化迁移，通常情况下会出现多个 VM 的配置都重复下发到一个物理接口上，很难做到针对每一个 VM 的精细化网络配置管理。

可行的办法只有先精细化区分流量（比如源 IP、源 MAC、VLAN 等），基于流量的识别再进行针对 VSI 的配置迁移。这样就又引入了新的复杂性和局限性。首先要求邻接物理交换机支持基于源 IP/源 MAC/VLAN 的 ACL 或流分类、以及带宽控制能力；其次要支持 ACL、MQC 的动态创建和删除能力，而在创建和删除时，很可能会影响其他 VM 的配置。

另外，当 VM 从一个物理服务器上迁移走之后，本地配置的去部署通常需要由用户预先配置好，此处若想实现智能化（由部署命令自动生成去部署命令），逻辑和实现也非常复杂。

Basic VEPA 方式也存在同样的问题，在拟定义的 802.1Qbg 标准中，也在考虑在 VDP 报文中增加附加过滤字段（比如 VLAN、源 MAC 等），来解决 VM 和物理接口 N:1 的问题。报文和设备处理的复杂性被放大了。

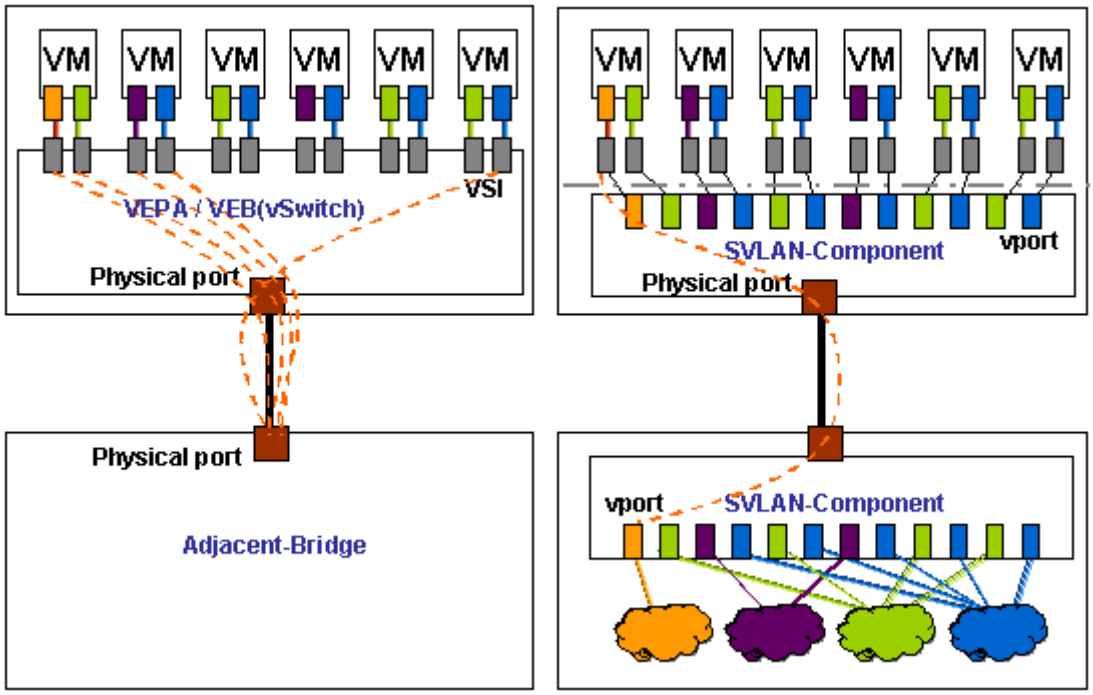

图12 VEPA/VEB、Multi-Channel/PE 物理端口映射对比

Multi-Channel、PE 方案的提出为此问题找到了最优的解决方法，即在邻接物理交换机出现了 vPort 的概念。这类逻辑虚接口可以实现和 VM 对应的 vNic/VSI 的 1：1 对应关系。

VM 迁移时，只需在对应的邻接物理交换机上动态创建一个 vPort（当然如果 VM 需要多个 vNic 时，对应也要创建多个 vPort），并将 VM 对应的网络配置 Profile 绑定到 vPort 上。不会对其他 vPort 产生影响。同时迁移前对应的邻接物理交换机只需要简单的将对应的 vPort 逻辑接口删除即可，不存在反向去部署的复杂性问题。

*注:目前邻接交换机对 Multi-Channel 的实现也可能不会创建 vPort 逻辑对象,而仅有针对 s-channel 的 sPort（可能直接对应一个 VSI，也可能对应一个 VEB 或 VEPA），这样其实和 VEM、VEPA 的处理类似，达不到 vPort 方式下的管理简化性。

四、数据中心虚拟化网络管理的下一个目标：链接即服务（CaaS）

针对上一节描述的网络“随需而动”的自动化能力的实现复杂性，目前正在形成标准的 VDP（VSI Discovery and Configuration Protocol ）方案对网络配置迁移方案进行了优化。

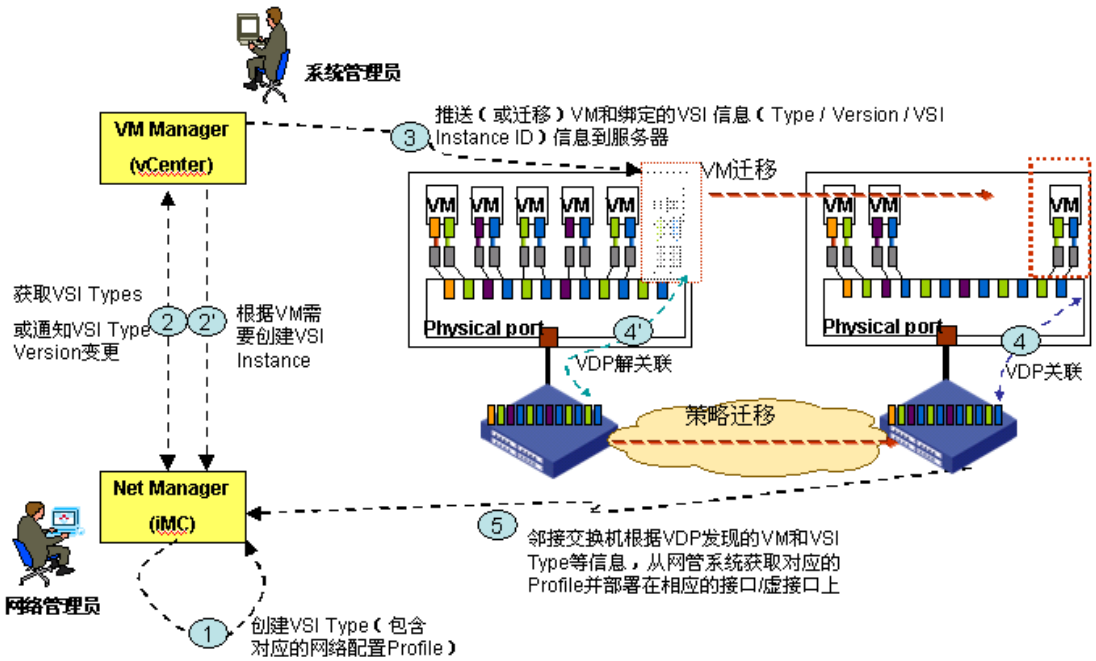

图13　支持VDP的虚拟网络配置迁移原理

① 针对不同的服务类型，网络管理员在网管系统中定义 VSI Type 以及其对应的网络配置 Profile；

②系统管理员在 VM Manager 中根据 VM 类型绑定合适的 VSI Type 并创建对应的 VSI 实例；

③系统管理员将 VM 和绑定的 VSI 信息（VSI Type、VSI Type Version、VSI 实例 ID）一同推送到支持虚拟化的服务器；

④服务器上的 VSI 在激活前，通过 VDP 通告给邻接网络交换机（包含 VSI Type 更新），邻接交换机就知道在哪个物理接口上需要链接什么类型（VSI Type）的虚拟机。

⑤邻接交换机使用该 VDP 发现的 VSI Type/Version 等信息向网管系统获取对应的 Profile 并部署到相应的接口上。同时，VM 迁移前的接入位置交换机也会通过 VDP 解关联通告，去部署相应的 profile 对应的配置。

在 VDP 方案中，当 vSwitch 为 VEB/Base VEPA 模式时，VDP 报文中需要通过附加的过滤字段区分出本 VSI（VLAN、源 MAC 等），并通告给邻接交换机，由邻接交换机在相应物理接口上根据上一节描述的基于流量识别进行 ACL、带宽等控制。

若 vSwitch 和邻接交换机为 Multi-Channel 或 PE 等支持 vPort 的模式，则不需要考虑区分 VSI，邻接交换机在发现 VDP 通告后，可以直接创建新的 vPort 并绑定从网管系统获取的 Profile；对应的拆除操作也仅仅是简单的删除 vPort 操作。

整体来看，引入 VDP 后，不依赖网管系统对 VM 接入物理网络的定位能力，提高了网络配置迁移的准确性和实时性。

关于系统管理员和网络管理员的职责划分问题，从这里可以看出，两者的分界线就在 VDP 所处的位置，对系统管理员来讲，只需要关注网络提供的虚拟服务器到邻接交换机的链接；相应的对网络管理员，则需要关注针对不同的应用系统（或 VSI 类型）应提供什么样的

网络接入配置。这样的配置抽象对系统管理员就是一种服务，针对链接的服务（CaaS，Connection as a service)。

五、结束语

企业采用虚拟化技术最初所追求的目标是通过建立横向和纵向的可伸缩、高弹性的服务器集群，优化资源的利用，提升服务水平，并降低管理开销。而如今在企业数据中心领域，在服务器虚拟化技术满足企业越来越多业务需求的同时，也对网络管理方式产生了日益显著的影响。虚拟化服务器、网络的融合管理已经成为当前数据中心 IaaS 能力的必要元素，各厂商都在相关领域进行尝试，也出现了不同的虚拟化软件厂商和网络厂商的联盟，不同的实现方案。另一方面，标准化工作也在快速进行，在完成统一的资源可视性、流量可视性，以及提高基于 VMware 的虚拟化网络配置迁移能力基础上，我们仍需紧跟标准，提供更为高效、可靠的“随需而动”的自动化管理能力。为企业打造业务永续、架构先进的企业数据中心。