化繁为简，推陈出新

运营商边缘网络虚拟化解决方案

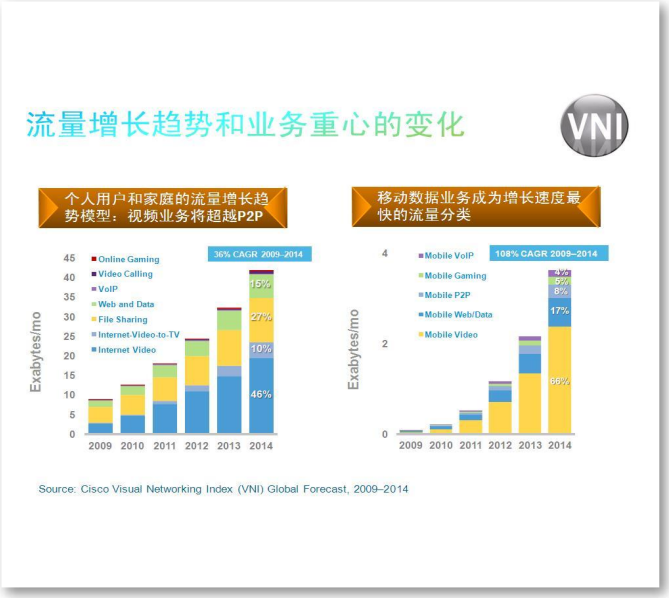
蔡德忠 (Dennis Cai)　思科边缘路由器产品线资深架构师
王伯剑　思科中国电信事业部客户解决方案架构师

目录

市场动态和驱动力

思科公司每年投入大量的企业资源和人力，通过与全球主流运营商、内容提供商、第三方市场研究机构以及各种国际标准化组织的合作，利用思科分布全球的、超过70%市场份额的路由器进行全球互联网流量统计和分析，发布面向全球及各大区域的互联网业务发展趋势的报告，思科称之为可视化网络指数VNI（Visual Networking Index）。这份报告根据最新的互联网流量数据，结合运营商业务发展形势，为用户提供了全面可信的流量增长模型预测，帮助运营商捕捉最新的互联网动态。

根据思科最新的VNI预测（见下图），未来的互联网无论从流量增长趋势还是业务发展重心上都会发生明显的变化，这些变化将最终影响到未来网络的流量规模、模型乃至网络架构的演进。从个人及家庭用户的互联网流量构成角度来看，视频业务将会在2012年全面超过P2P业务；从流量模型的角度来看，更多的流量将会从目前由用户提供转向由数据中心提供，流量聚合的趋势愈发明显。此外，移动数据业务随着Wifi、3G和LTE的大量部署将快速增长，成为流量增速最快的业务类别，这也将对今天运营商们的移动回传网络和核心网产生巨大的影响。例如，目前在全球范围内3G和LTE发展比较成熟的运营商都开始使用IP-RAN技术替代传统的MSTP建设移动回传网络。



今天的互联网已经进入了“内容为王”的时代，在这种商业模式下，互联网的生态环境在向内容提供商倾斜，这一点可通过最新的全球互联网流量分布统计清晰地看到，根据第三方研究机构发布的报告，互联网整体的流量分布和贡献比率在过去几年发生了巨大的变化，第一个变化是流量更加聚合，2007年全球排名前十名的流量贡献者占据了整个互联网30%的流量，而到2009~2010年，这一比例上升到了40%，流量向内容提供商聚合的效应愈发明显；第二个变化是流量贡献者的排名在从电信运营商向内容提供商转移。2009~2010年，Google和Comcast（全美最大的有线电视提供商）首次进入了前十名的名单。这些数据的变化都印证了当今内容提供商在互联网生态系统中占据更大话语权的事实。但是伴随互联网产业规模和流量的快速增长，运营商数据业务营收的增长比率却大大落后，与此同时为了应对流量增长的压力，网络基础设施的投资和运维成本却逐年增长。运营商面对的是增量不增收的困境。

为了突破这样的业务困境，全球的运营商们都在积极地向体验提供商转型，这种转型的关键之处在于开源节流，一方面，通过互联网内容和应用的经营和服务，提高业务收入、加强差异化竞争，广开源头；另一方面，简化和优化网络基础架构、降低运维复杂度和建网成本，最终节省每字节流量的开销。作为IP技术的引领者和倡导者，思科一直致力于IP技术的发展和创新，并通过思科领先的产品及解决方案， 帮助运营商解决业务发展过程中所遇到的各种挑战。

本文所介绍的网络虚拟化技术，就是思科针对运营商边缘网络所研发的一项创新解决方案。利用这一解决方案，运营商不仅可以提高网络的可靠性及健壮性，极大改善用户对服务永续的体验需求；更为重要的是可以简化网络的架构、设计及部署，使得网络的运维成本大幅降低，根据第三方调查咨询公司ACG所进行的商业模型分析（见附录），采用思科此项创新技术，可以帮助运营商节省高达70%的运维成本。

运营商面临的挑战和网络虚拟化诉求

在IP和IT技术发展的20多年历史中，虚拟化并不是一个新鲜的话题，为了解决资源共享、安全隔离、管理维护等不同方面的需求，从路由交换、安全产品、存储以至服务器，从二层交换、三层路由直至应用层，出现了各种不同的虚拟化技术，按照虚拟化方式的不同，大致可分成两类：“一虚多型”和“多虚一型”。前者的目的是将资源细分、提供多用户使用的环境、并且保证安全隔离，比如：MPLS/VPN技术、虚拟路由器技术、VLAN技术、虚拟桌面等都是属于此类技术范畴，而后者的目的多是将资源进一步整合，提供更高的交换容量和可靠性，简化和优化网络架构，消除不必要的网络协议部署，使得网络更加简单易用，比如：堆叠技术、核心路由器的集群技术、数据中心的VSS技术等都是属于此类技术范畴。

本文所介绍的思科运营商边缘网络虚拟化解决方案即属于后一种类型，具体来说该解决方案实际由两种技术组成：

- 虚拟集群技术，也称作Clustering技术。简单地说，就是将两台路由器虚拟成逻辑上为单台的超大型路由器，两台路由器在控制和管理平面完全统一，运维人员如同管理一台设备一样实现对虚拟集群的配置和管理。
- 接入虚拟化技术，也称作Satellite技术。在接入虚拟化技术中，边缘路由器所连接的扩展设备都被虚拟化成路由器的一张板卡，就好像围绕在地球周围的人造卫星一样，这些成为“卫星”的扩展设备的所有业务配置及管理都统一交由边缘路由器负责，无需再单独配置管理接口，逐一进行配置下发和维护。

在进行更深入的技术细节讨论之前，有必要结合运营商网络的实际架构及运营商在网络设计和运维过程中所遇到的实际问题，来分析当前运营商在边缘网络所面对的主要挑战和需求，这也将有助于理解思科边缘网络虚拟化技术设计的初衷和应用的场景。

边缘路由器的用户状态热备份

在当前运营商网络架构中，边缘路由器无论是承担骨干网PE的角色，还是作为城域网的BRAS和SR，都需要直接终结用户的业务。作为用户接入网络的第一跳网关，边缘路由器上维护了用户相关的业务属性、配置及状态，如：用户的IP地址、路由寻址的邻接表、DHCP地址绑定表、组播加入状态、PPPoE/IPoE会话、Qos和ACL属性等等，这些重要的表项和属性直接关系到用户的服务质量和体验。因此，为实现“业务永续”的目标，作为消费者的最终用户期望运营商提供高冗余的网关保护，这一点对于企业客户或高端个人客户尤为重要，即：所有与用户相关的业务属性、配置和状态信息都需要在冗余网关之间进行实时备份，当其中一台边缘路由器出现故障而脱网之后，另外一台冗余的边缘路由器能够即时替代原有设备投入运行，用户信息不因故障而丢失，无需重建用户会话和重新下发用户属性，真正做到对最终用户的无感知切换。

运营商一直在寻求实现用户状态热备份的解决方案，传统的做法是采用协议扩展的方式。如：利用VRRP实现双机热备份，但是标准的VRRP协议无法对用户状态信息进行备份，因此出现了很多厂家私有的协议扩展，IETF标准组织也开发了ICC框架来解决这种跨机箱备份的需求。但是这种基于协议扩展的方式存在一些局限性：第一，为实现可靠的状态备份，这种扩展协议通常要增加确保可靠性的机制，因此协议通常较复杂；第二，状态备份需要实时进行，对于大用户量并发备份需求，协议扩展方式将增加设备在控制平面的负担，从而影响设备的整体性能；第三，最大的局限性来自可扩展性差，基于协议扩展的方式经常要随着变化的功能需求而不断进行修改，这导致了运营商网络的频繁升级和潜在的安全隐患。

因此，为了避免上述局限性，思科独辟蹊径，采取了完全不同的方式设计虚拟集群技术，该技术汲取了思科在集群技术方面领先的技术精华，从系统架构的高度在根本上解决了用户状态热备份的需求。

网络的高可靠性和快速收敛

今天运营商的很多关键业务都是通过IP网络进行承载和传输，如：IMS业务、移动回传、大客户专线业务、R4移动核心网承载等等。对于这些关键业务而言，网络的高可靠性是衡量网络承载质量的重要指标，因此很多大型的运营商网络从规划建设之初，就将如何实现网络的高可靠性，以及出现故障后如何提供快速收敛作为重中之重而优先考虑。因此，从网络技术角度出现了：TE/FRR、IGP快速收敛、IP/FRR、VPN FRR、PWE3 冗余、BGP NHT、Unique RD、BGP PIC、VRRP、链路捆绑、REP、Flexlink等等不胜枚举的各种故障保护和收敛技术。如何针对不同的故障场景，选择和部署这些技术以及如何在出现故障时进行诊断成为了运营商规划设计以及网络运维人员最为头痛的问题之一。

思科的虚拟集群技术通过多虚一的方式，将冗余的网络边缘路由器虚拟成逻辑的单台设备，不仅减少了网络中网元数量、降低了IGP的路由规模，而且使得上述提到的绝大多数收敛保护技术变成不必要的选项，在极大改善网络收敛效果的基础上（从几百毫秒到小于50毫秒），大大简化了网络设计、降低了网络运维的人力和成本。

庞大的接入汇聚网络管理

全球运营商们都在为提高用户接入带宽、改善用户接入条件而进行努力。在中国，运营商们进行的“光进铜退”运动，将用户的平均接入带宽由2Mbps提高到20Mbps。随着每用户接入带宽的提高和大量xPON设备的部署，城域网在接入和汇聚层面的交换机数量大幅提高。根据测算，在一个进行了“光进铜退”改造的中等规模城域网中（80~100万互联网用户数），其接入和汇聚层面的交换机数量有2000~3000台的规模。这些接入汇聚交换机大多分布在无人值守的远端模块局和用户侧机房，功能要求和配置非常简单，除特殊应用场景， 绝大多数情况下端口都是超卖且上联无冗余保护。为了实现对庞大的接入汇聚网设备的管理，运营商不仅需要保持一定规模的基层运维团队，而且还要为每台设备配置单独的管理端口和IP地址，甚至需要购买单独的网元许可（License）以实现集中网管平台的远程配置下发和业务开通。一旦出现接入汇聚网络的故障，诊断和故障恢复周期都相对较长。

思科的接入虚拟化技术通过协议扩展和IPC机制（进程间通讯），将连接到边缘路由器上的数以千计的扩展设备，虚拟成边缘路由器上的板卡，网管人员可以如同配置边缘路由器一样，通过对扩展板

卡的操作，实现对这些“卫星”设备的统一配置和业务开通，甚至进行批量的软件升级，整个网络虚拟化成为单独的网元。这些数以千计的扩展设备都被看做是边缘路由器的延伸板卡，边缘路由器和扩展设备之间可以光纤直驱，也可以穿越城域以太网、MSTP或者WDM波分网络，思科的接入虚拟化技术能够延伸至城域网中任意的最末端网络位置。而且，扩展设备可以以星型、环型、跨机箱链路捆绑等方式连接到边缘路由器上，这使得整个接入汇聚网的拓扑更加灵活和可靠，适合承载各类关键的业务。

目标网络架构

思科的边缘网络虚拟化技术将运营商网络架构“化繁为简”，下图对比了采用思科边缘虚拟化技术前后的网络架构，可以看到在原有的网络架构中，OLT设备通过汇聚交换机层层级联，最终双归连接到冗余的BRAS或SR。当采用了思科边缘网络虚拟化解决方案之后，思科虚拟集群替代了冗余的BRAS和SR，扩展设备替代了汇聚交换机并一直延伸至OLT，所有的扩展设备都虚拟化成集群上的板卡，消除二层网络STP，网络架构更加扁平和简单。

图1 采用网络虚拟化技术之前

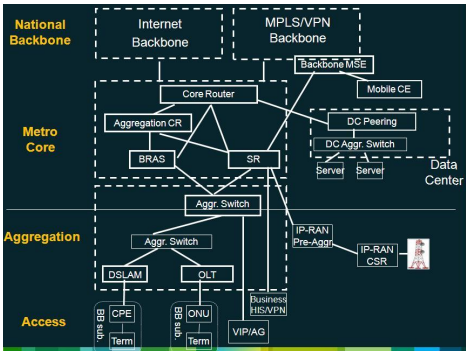
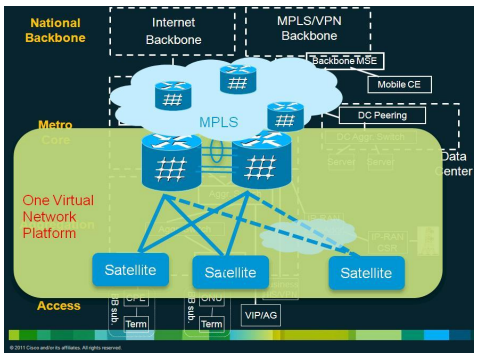


图2 采用网络虚拟化技术之后



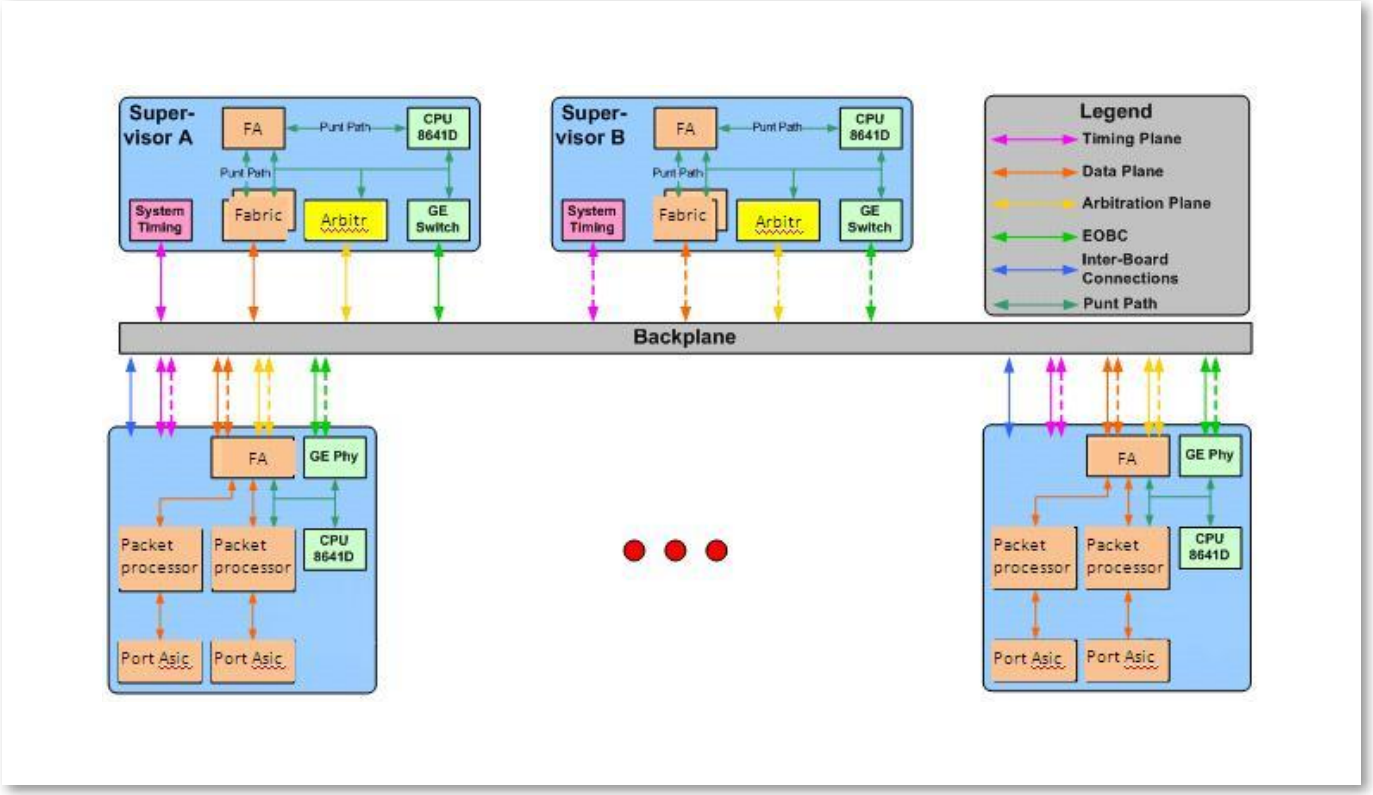
思科虚拟集群技术

思科在2004年推出业内最为先进的核心路由器CRS时，创新地提出了集群路由器的概念，即将传统分布式路由系统的单级交换架构改变成基于BENES的三级交换架构，因此路由器的交换矩阵得以与线路卡物理分离，通过不断增加线路卡机框，就可以不断提升整个路由系统的容量和端口密度，而且路由系统在逻辑上始终保持单一。集群技术的出现在当年彻底颠覆了人们心目中传统路由器的形态和演进的方向，并且时至今日一直成为网络架构扁平化的核心技术推动力之一。此后，业内形成普遍共识：核心路由器必须支持集群技术。

今天思科将在集群技术方面的领先优势和技术积累延伸至运营商网络边缘。通过虚拟集群技术将冗余的边缘路由器虚拟化成逻辑上单一的超大规模路由器系统。不仅从架构上解决了跨机框的用户状态备份的需求，而且从逻辑上将网络的规模和复杂度降低了一倍，网络边缘的保护和收敛问题得以大大简化，本章节中将重点讨论虚拟集群技术的技术实现。

虚拟集群的原理

在介绍思科虚拟集群技术之前，首先来回顾典型的分布式路由系统的架构图，如下图所示，分布式路由系统的典型特征就是：管理、控制与转发平面是完全分离的，管理平面和控制平面主要驻留在路由引擎，而转发平面则驻留在线路板卡。分布式路由系统通过专有的基于以太网的带外管理通道EOBC（Ethernet Out-of-Band Channel）在集中式的路由引擎和分布式的线路板卡之间同步各种管理和控制平面的状态信息，使得各个组件保持严格的同步，以避免可能出现的路由黑洞或其它方面的异常。由此可见，EOBC的可靠性和效率对于分布式路由系统是至关重要的。因此，通常而言，EOBC在分布式系统中都是专有的通道，不会与交换矩阵通道进行混用。同时，为了实现高可靠性，线路板卡与主备路由引擎之间存在冗余EOBC连接。一旦引擎切换或者故障，备份的EOBC通道能立刻投入正常工作。



思科虚拟集群技术的关键创新之一就是将EOBC延伸至单机系统之外，通过延伸的EOBC通道，将物理上分离的两台单机系统，从逻辑上（控制和转发平面的角度）构成一个新的虚拟集群系统，在新的集群系统中：

- 管理平面统一，用户可以通过单一的管理接口去访问、配置和控制分布在两个物理单机系统上的所有硬件资源（引擎、板卡、电源和风扇系统等）；并且维护单一的系统配置文件。

- 控制平面统一，新的虚拟集群系统对外表现为单一的路由节点，所有的网络节点只需与虚拟集群系统建立邻居关系即可，控制平面的邻居数量降低一倍，也不再需要VRRP/HSRP等路由冗余保护协议。虚拟集群系统获取的控制平面消息经过主路由引擎计算和优选后，通过EOBC下发至集群内的所有单机系统的板卡上。如下图所示，思科的边缘路由器上提供对外的EOBC接口（标准SFP+接口），用户通过光纤（单模或多模依赖于SFP+光模块的选择，虚拟集群技术对此无硬件限制）即可将两台单机系统进行连接，为了保证EOBC连接的高可靠性，思科边缘路由器的每个路由引擎卡上都提供了冗余的EOBC接口，EOBC接口在单机系统之间两两互联，内部运行思科快速收敛协议防止EOBC连接出现环路，并能够在小于十几毫秒内实现EOBC链路的切换。任何情况下，仅有一条EOBC链路在工作状态。在实际部署中，EOBC的连接无需用户配置，完全即插即用。

- 转发平面独立，虚拟集群中两台物理单机系统尽管在管理和控制平面上统一，但是在转发平面上是保持独立的，即两台单机系统都可同时转发流量，使得集群系统的整机交换容量提升一倍。在边缘路由器部署的典型场景下，以南北向流量为主（这点与核心路由器所面对的网状流量的场景不同）。正常情况下，特别是接入节点以双归方式连接边缘路由器时，通常在两台设备之间有少量穿越流量，传统的网络设计如图3所示，两台设备之间存在互联链路：一方面用于故障保护，另一方面用于出现流量不对称时提供流量绕转能力；当部署虚拟集群技术之后（如图4所示），从转发平面的角度，虚拟集群内的两台单机系统之间仍然会保留互联链路。因此，当运营商从传统的网络设计迁移到虚拟集群设计时，网络拓扑连接和端口链路需求几乎完全相同，可以实现平滑迁移。

图3 传统边缘路由网络设计

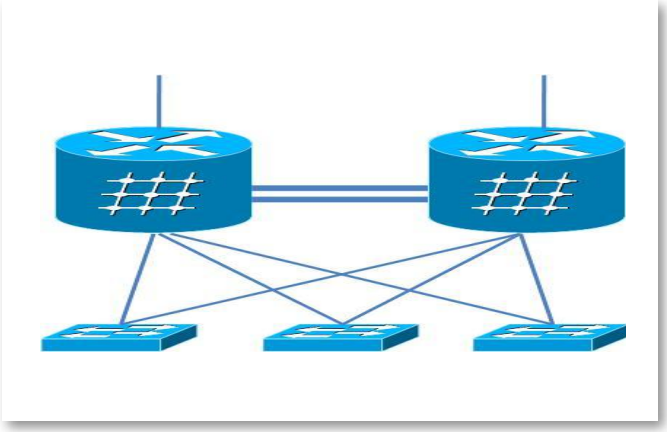
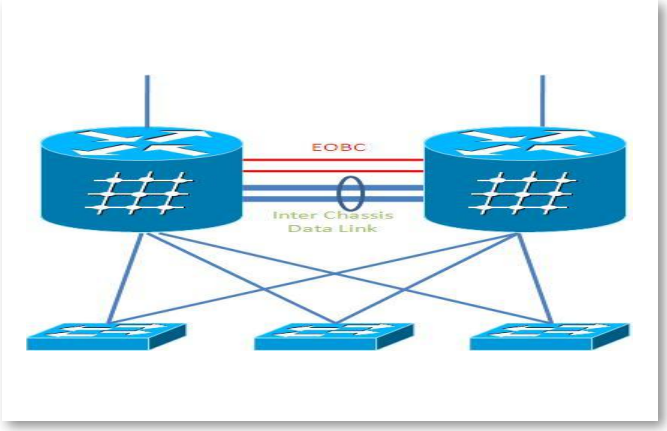


图4 思科虚拟集群的连接示意图



思科的虚拟集群技术除了将EOBC扩展至机箱之外，还在软件操作系统上针对EOBC同步进行了专门的优化和加固，改进了EOBC的同步效率、保证通讯可靠性的确认机制、EOBC故障检测机制、失效重传机制等；一系列软件方面的开发提高了EOBC的容错性，也增强了系统对EOBC消息延时的容忍度。根据思科内部的测试，即便EOBC消息的延时达到100ms左右（从美国东海岸到西海岸的网络传输单向延时大约30ms），思科的运营商操作系统IOS-XR仍然能够保证EOBC消息的正确性。这些软件方面的优化，提高了虚拟集群技术部署的灵活度。集群内的两台单机系统之间不再有物理距离的限制，只要EOBC的光纤可达，或者中间通过波分系统、交换机等，都可构成虚拟集群系统。这使得虚拟集群技术既可在同机房部署，又可跨机房部署。

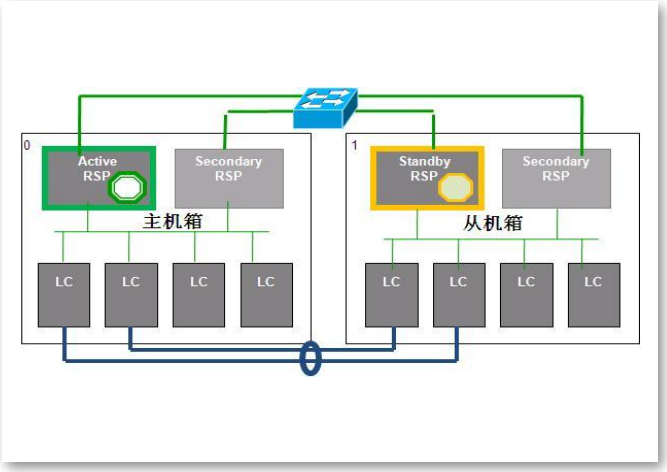
虚拟集群的控制和转发平面

组成虚拟集群的两台单机系统都存在独立的冗余路由控制引擎和交换矩阵，为了确保虚拟集群能够有效地在两台单机系统之间进行控制平面状态备份，以及优化流量的交换效率，虚拟集群在分布式路由系统的实现机制上进行了一些改进：

首先，从控制平面角度。整个虚拟集群内：单机系统分成主机箱和从机箱；路由控制引擎分成：主引擎（Active）、备引擎（Standby）和辅助引擎（Secondary），思科虚拟集群技术支持四个路由控制引擎的全冗余配置模式（每个单机配备两个路由控制引擎）。在系统启动/初始化以及EOBC重新连接时，虚拟集群内将进行控制权的选择，其基本原则是：按照单机系统启动的顺序，率先完成启动过程的单机被选为主机箱，也可以通过命令行配置进行指定。主机箱的主引擎被选做整个虚拟集群的主引擎（Active），从机箱的主引擎被选作整个虚拟集群的备引擎（Standby），其它引擎都为辅助引擎（Secondary），仅在主备引擎都失效时才替代工作。在任何时候，整个集群内有且只有一个引擎是主用状态（Active），另外一个引擎处于热备份状态（Standby），因此从简化设备需求的角度，虚拟集群也支持在每台单机系统内仅配置一块路由控制引擎的场景。

这种设计使得虚拟集群能够如同单机分布式路由系统一样，在主备引擎之间进行控制平面状态和信息的实时备份，因此集群系统下的SSO（基于状态切换）/ NSF（无中断转发））/ NSR（无中断路由）等操作与单机系统没有差别，扩展的EOBC将两台单机从控制平面角度完全同步在一起。这种实现方式最突出的优势在于，从系统架构的高度彻底解决了跨机箱热备份问题，单机系统中已支持热备份的控制协议可以从第一天起就在虚拟集群系统上提供跨机箱热备份，这使得运营商级的高可靠性需求通过虚拟集群技术得以快速实现。

图表5 虚拟集群的控制平面示意图



其次，从转发平面角度。所有数据报文在转发时都优选本机直连的邻接表信息，尽量避免跨越单机的横向流量。对于流量负载分担的需求，虚拟集群技术推荐双连接拓扑，即用户CPE以双归方式连接到虚拟集群的两台单机，同时虚拟集群的两台单机也分别连接到核心路由器。正常情况下，集群内单机之间的背靠背链路捆绑组上无用户流量经过（特别是单播流量），只有当出现单连接时，背靠背链路捆绑组才用于提供流量的绕转。例外情况是组播应用的场景，由于组播的复制点可能跨越集群的不同机箱，因此背靠背链路捆绑组也被用于提供跨机箱的组播复制。在思科的分布式路由系统中，组播的复制分成两级：一级是交换矩阵复制，一级是出接口板卡复制。对应于虚拟集群系统，单机之间的背靠背链路捆绑组模拟了交换矩阵的行为，因此只有一份组播复制流量通过背靠背链路捆绑组。思科的虚拟集群技术并没有改变分布式路由系统的转发架构，两级转发（Dual Stage forwarding）和两级组播复制等行为仍然与单机系统完全一致。

虚拟集群的高可用性保证

由于虚拟集群技术将两台物理上分离的单机系统进行了逻辑上的虚拟化，并且两台单机系统还可以跨机房部署，因此确保虚拟集群的高可用性成为该技术在实际部署时，需要着重考虑的问题，思科主要从以下几个方面来提高虚拟集群的可用性：

第一，思科虚拟集群技术在整个集群系统内最多支持四个路由控制引擎，除了主备引擎之外，每个单机系统中还存在辅助引擎提供备份功能。这为整个集群系统带来了防御路由引擎二次故障的保护，相比较单机系统具有更高的可用性。

第二，思科在单机系统之间的EOBC通道上提供了检测机制，利用快速UDLD（Unidirectional link detection）协议，每间隔50ms互相发送hello报文，连续一定数量的Hello报文没有收到，EOBC通道就切换至备份连接，因此可以在几百毫秒内检测到单机系统的故障。

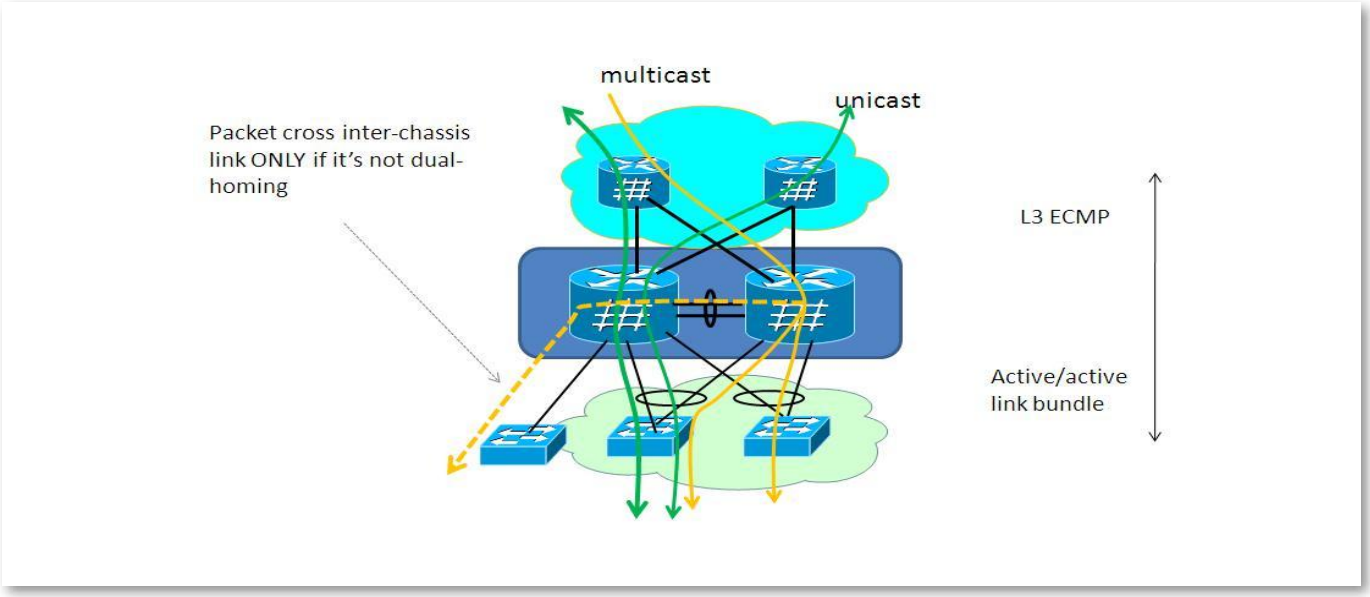
第三，虚拟集群内两台单机系统共享相同的配置文件，因此当单机之间的所有EOBC通道都中断后，集群会分裂成两台独立的单机系统，由于这两台单机系统的配置相同，这将会引起网络控制平面的混乱（比如：由于两个单机系统配置有相同的RouterID）。因此，在这种情况下，集群系统中的原从机箱会自动进入板卡端口关闭状态，此时只有EOBC接口仍保持激活，直至EOBC连接恢复后，从机箱重新启动可再次加入集群系统中。

虚拟集群的优势与比较

思科虚拟集群技术的优势可简要地概括为三个词：更大、更快和更简单：

- 虚拟集群提供双倍的单机交换容量。
- 部署虚拟集群技术之后，网络的快速收敛和高可靠性设计因为节点的虚拟化（两台设备变成一台设备）而简化，边缘路由器节点保护这一在传统网络路由设计中非常复杂的问题，在虚拟集群场景下不复存在，运营商无需花费大量精力进行复杂的网络设计和维护。

图表6 虚拟集群技术的转发平面示意图



图表 7 虚拟集群技术与其他传统提供跨机箱保护技术的对比

	虚拟集群	跨机箱链路捆绑 (MC-LAG)	冗余网关备份协议扩展
管理平面	一个	两个	两个
控制平面	一个	两个	两个
配置	简单，即插即用	复杂	复杂
系统交换容量	增加一倍	利用VLAN负载分担可实现两台单机同时工作	利用VLAN负载分担可实现两台单机同时工作
接入链路状态	Active/Active接入	Active/Standby接入	Active/Standby接入
跨机热备份复杂度	简单	仅实现少量协议跨机备份，如linkbundle，IGMP Snooping	复杂，支持PPPoE/IPoE，需要随协议增加而随时扩展
网络收敛	<50ms	支持L2vpn，秒级	仍需要部署复杂的快速收敛技术，秒~分钟级
跨局址部署	可以	可以	有局限性

- 在传统网络架构中，尽管利用了各种复杂的快速收敛技术，节点保护的收敛时间仍然在几百毫秒至几秒之间（依赖业务部署的规模和类型）。而利用思科虚拟集群技术，由于单机的故障不影响整个集群的控制平面状态，因此收敛过程变成本地行为。在思科所进行的内部测试中，即便叠加了一定规模的L2vpn/L3vpn等业务，虚拟集群的收敛时间仍然在50毫秒之内，与业务叠加的规模和类型无关（Service Convergence Independence）。
- 部署方式灵活，即可同址部署又可跨机房部署。
- 跨机箱状态热备份在系统架构上自动实现。

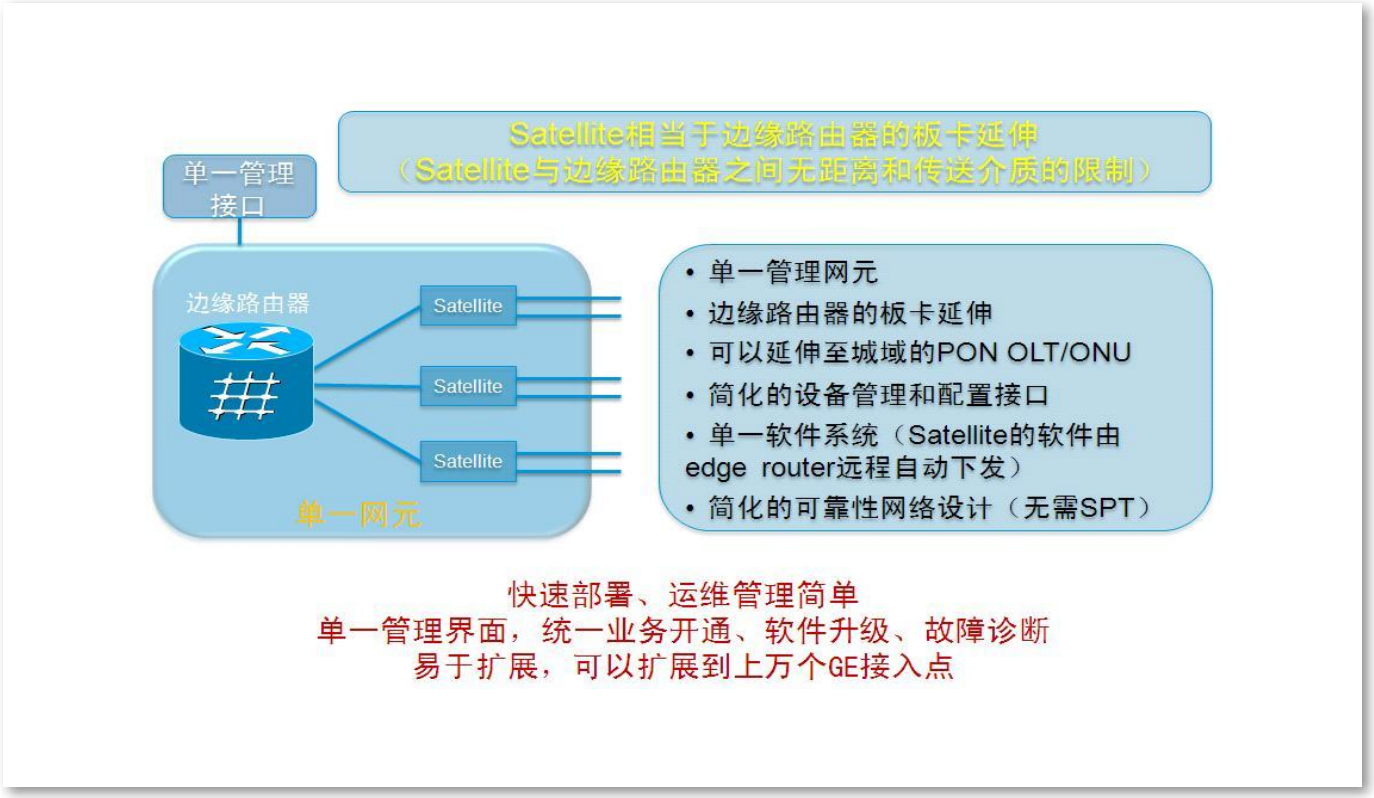
思科接入虚拟化技术

随着网络流量的快速增长，路由交换设备的容量越来越大，运营商对设备的高速以太网端口的密度需求也越来越高。今天的网络边缘路由器已经开始进入200G平台时代，每个槽位的物理板卡可以提供400Gbps以上的吞吐容量。由于板卡物理尺寸的限制，在如此高带宽的槽位上无法提供与之相对应的满配置线速千兆以太网端口，端口类型的灵活性上也受到一定的局限。此外，运营商的“光进铜退”使得xPON在网络接入层广泛部署，千兆/万兆以太网端口的汇聚在城域网中非常普遍，因此运营商不得不部署大量的以太网交换机来满足此需求。这种做法，一方面增加了网络的层次，带来了运营商的额外投资；另一方面，接入／汇聚网中众多以太网交换机的管理、配置和维护，以及复杂的二层网络设计都成为运营商网络维护成本居高不下的主要原因。

因此，思科在规划未来网络架构发展方向时，针对运营商降低OPEX成本和提供更加扁平化、更加简单的网络架构的诉求，设计开发了网络接入虚拟化的解决方案。利用此方案，运营商可以基于大容量的边缘路由器迅速为用户提供成千上万的以太网端口，降低了端局/模块局机房的需求，符合运营商“大容量、少局址”的设计理念。并且所有的以太网端口的管理、配置和维护都可通过统一的网管界面一次性地批量下发，这种PnP“即插即用”的特性加快了业务的开通效率。同时，接入虚拟化解决方案也简化甚至消除了网络的汇聚层，二层网络设计中很多复杂的问题，如：环路防止、泛洪、收敛保护等都不复存在，运营商的维护成本将大幅降低，本章节中将重点讨论接入虚拟化的技术实现。

接入虚拟化的概述

思科接入虚拟化解决方案包括两个组件，一个是实现集中控制的边缘路由器，也被称作中心节点；另外一个是提供端口扩展和远程接入的扩展设备，也被称作远端节点。所有的远端节点都通过标准的万兆以太网接口（SFP+）连接到中心节点上的高密度万兆接口卡。当远端节点完成初始化注册和配置之后，就被虚拟化成中心节点上增加的一块扩展板卡，而后网络运维人员就可以如同配置边缘路由器一样直接实现对所有远端节点端口的统一管理和配置。远端节点与中心节点之间既可以通过光纤直驱连接（当采用单模SFP+时，光纤直驱的距离可达80KM），又可以透过城域以太网、MSTP或者WDM传输网络进行连接。从距离的角度，思科接入虚拟化解决方案可以将远端节点延伸至城域网的最末端机房、甚至是客户侧网络位置。远端节点上还能够提供丰富的用户侧接口类型：千兆光



口/电口、百兆光口/电口、xPON接口等，这使得远端节点可以应用到更广泛的用户场景。

从软件架构的角度，所有的配置和功能对于远端节点和中心节点是完全一致的，甚至远端节点本身的软件包都可直接从中心节点下载升级，配置和管理变的简单而直接。

接入虚拟化的控制和转发平面

为了实现远端节点与中心节点之间控制平面的统一，思科开发了用于内部通讯的轻量级协议SDCP（Satellite Discovery and Control Protocol），SDCP协议主要包含两部分功能：第一，用于远端节点的自动发现和注册，类似于CDP协议的链路层心跳消息在远端节点和中心节点之间定期快速发送，同时也可以用来检测两者之间可能的链路故障。为了标准化和互通性，思科计划未来基于BFD协议实现此功能；第二，用于远端节点与中心节点之间的进程间通讯、控制平面的状态同步、配置下发以及软件版本的远程升级；为了确保通讯的可靠性，这部分功能利用基于TCP层的协议消息实现。

从转发平面的角度，远端节点上的接口被分成接入端口和上联端口两种类型，接入端口之间自动进行安全隔离，只保留接入端口与上联端口之间的本地交换功能。远端节点接收到的所有接入侧流量都统一上送到中心节点进行转发处理，为了在转发流量中标识出不同的远端节点以及远端节点的不同接入端口，思科在中心节点与远端节点之间的数据包上增加了一层标签进行标识，因此当中心节点接收到远端节点的流量后，能够通过该标签识别出流量的准确来源。

目前思科正在推动中心节点与远端节点之间控制协议的标准化，未来将会纳入到IEEE 802.1Qbh标准体系中，实现与其它基于IEEE 802.1Qbh的设备的互联。

业务部署和软件升级

远端节点上的每个接入端口都以独立端口形式在中心节点上直接配置，从管理和部署的角度，运维人员觉察不出远端节点的端口与中心节点的物理端口之间的差别：命令行配置、端口统计、SNMP消息、故障调试等完全一样。事实上从软件架构的角度，中心节点和远端节点在功能特性上完全一致。绝大部分控制和转发平面

的处理都在中心节点上完成（比如：组播的协议处理、路由表计算、G.8032、链路捆绑等），尽量简化远端节点，但是远端节点上也支持一些需要端到端部署的功能（如：以太网OAM、组播Snooping、1588v2等）。对于Qos的部署，用户只需在中心节点上对于远端节点的接入端口配置Qos策略，IOS-XR软件就会自动从远端节点接入端口到中心节点端口之间逐级启动相应的Qos功能，包括远端节点接入端口的限速、上联端口的流量整形和队列、中心节点端口的限速、甚至是四级的层次化队列调度。

为了更大程度上体现接入虚拟化易于管理和维护的优势，思科的远端节点可以选择自动或按需两种方式通过中心节点统一进行软件升级，远端节点的相关软件以单独软件包的形式发布，通过中心节点运行的IOS-XR操作系统，以Pie文件的方式安装和卸载，既保证了远端节点软件开发的灵活性，又确保了与中心节点的IOS-XR软件的一致性。

接入虚拟化的拓扑和保护

运营商接入/汇聚层按照光纤的分布主要分成链状、树型以及环型，为了提供高可靠的接入连接，通常以双上行连接和环型拓扑为主，思科的虚拟化接入解决方案可以适应各种接入拓扑场景，利用链路捆绑、环路收敛协议等实现保护（如右图）。

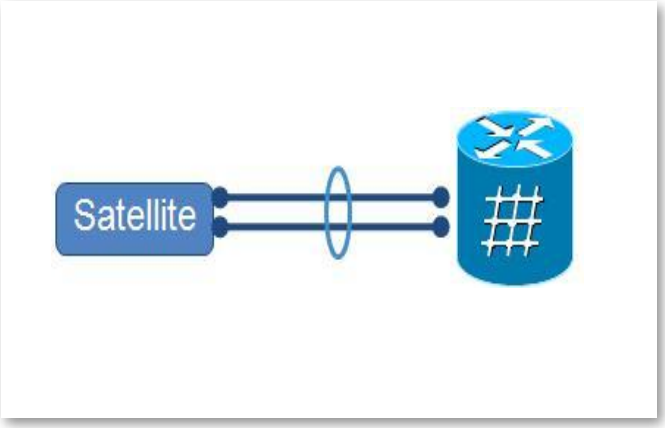
接入虚拟化的优势

思科的网络虚拟化解决方案具有很好的可扩展性和灵活性，可以支持多种形态的扩展设备作为远端节点，比如：高密度千兆以太网端口的扩展板卡、基站路由器、OLT、第三方交换机等。

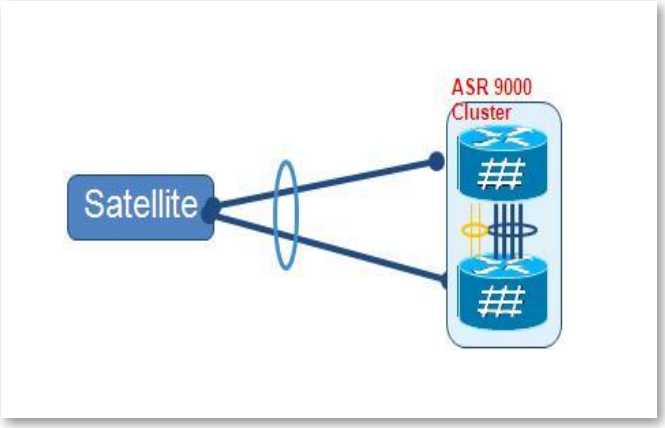
思科接入虚拟化技术的优势可简要地概括为三个词：更密集、更简单、更易用。

- 远端节点相当于边缘路由器大容量板卡的远端延伸，运营商可以快速提供成千上万的以太网和xPON端口。
- 统一的管理和配置接口，单一网元，一致的软件特性，运营商的业务开通更加简单。
- 即插即用、远程软件升级、批量配置下发、简化的网络保护方案，降低运营商运维成本。

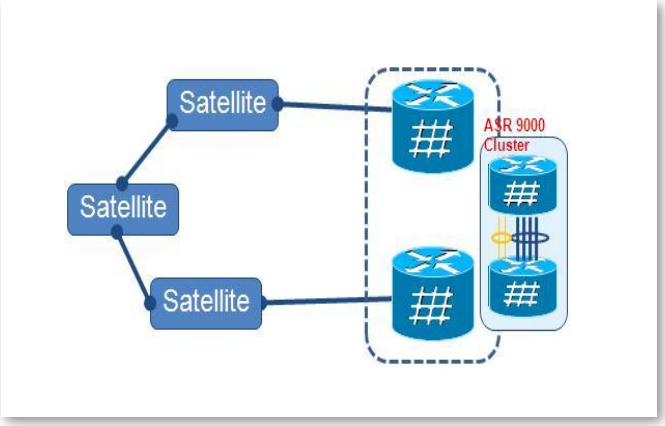
图表8 链路捆绑双上行方式



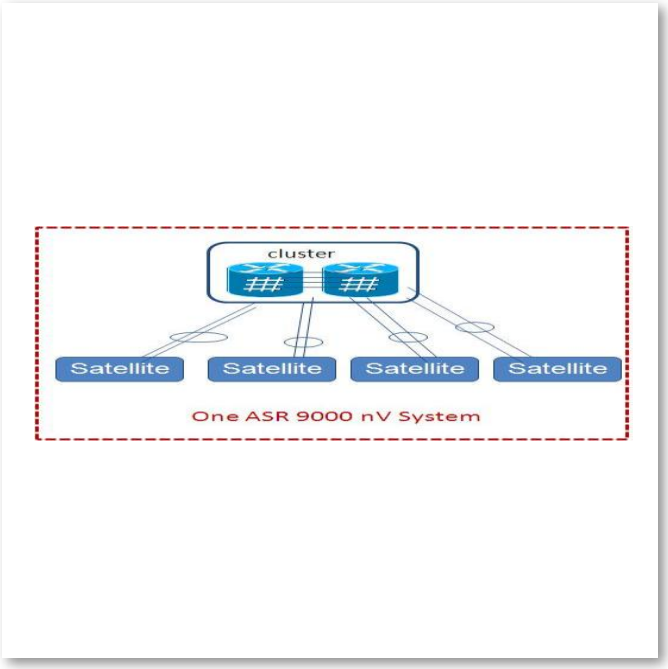
图表9 跨虚拟集群双上联方式



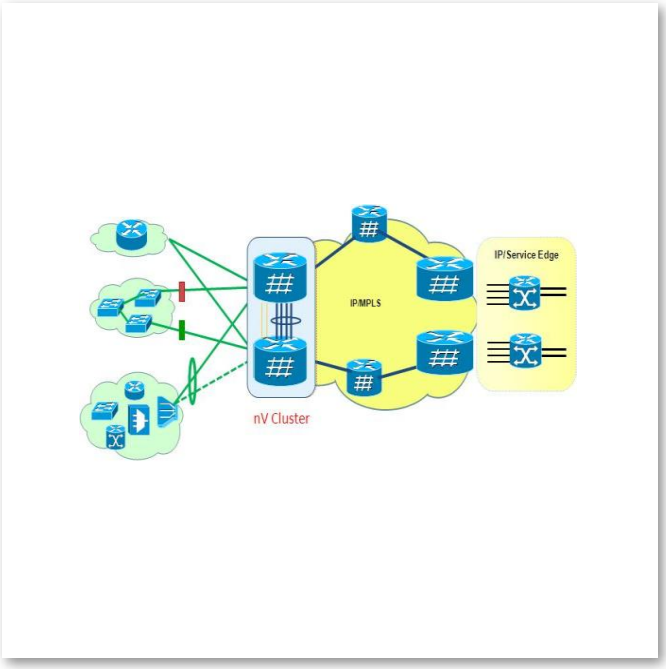
图表10 环形连接



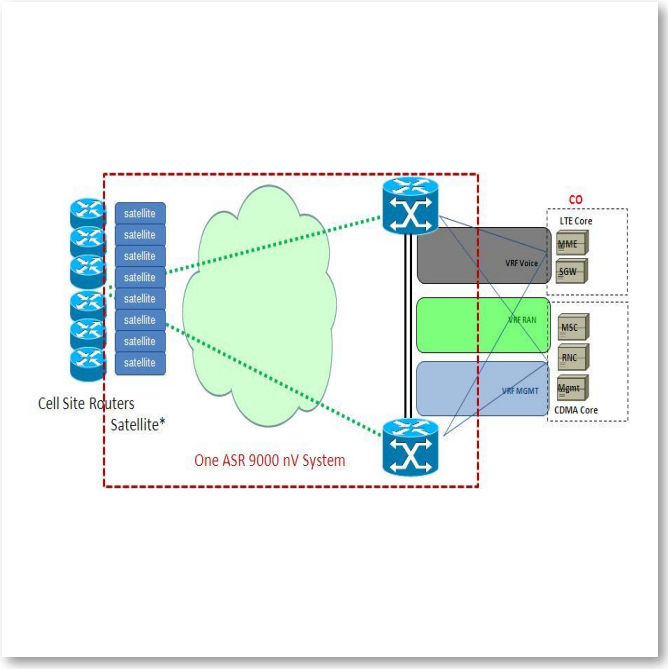
图表11 FTTX的宽带用户汇聚



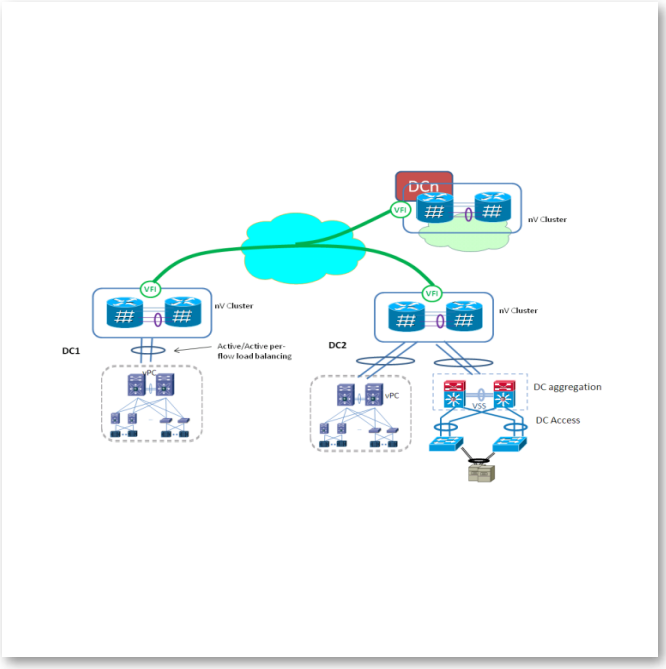
图表12 大客户的专线及VPN接入



图表13 移动回传网络



图表14 数据中心互联



思科网络虚拟化技术的应用场景

思科的网络虚拟化解决方案通过进一步的产品开发，以及与运营商的合作，可以应用到多种不同的业务场景。例如，通过支持OLT交换机，可以应用到FTTX宽带用户的接入；通过支持基站路由器，可以应用到移动业务的回传等。

总结

通过以上的分析和介绍，可以看到思科的网络边缘虚拟化解决方案确实是IP NGN领域针对未来网络架构演进的技术创新，该解决方案源自思科对运营商业务、管理和运维需求的深刻理解、并着眼于未来流量增长和新业务发展的愿景，切合运营商提高业务体验、改善运营效率、降低运维成本的现实压力，适合运营商广泛部署。

附录

- 思科VNI互联网业务与流量预测报告：http://www.cisco.com/en/US/partner/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html
- 思科下一代IP NGN边缘网络产品及解决方案：http://www.cisco.com/web/solutions/sp/asr9000.html
- 第三方研究机构ACG建设高扩展网络的商业模型案例分析：http://acgresearch.net/WebTop/download.aspx?m=CiscoASR9000v17.pdf
- IETF的ICC架构：http://tools.ietf.org/html/draft-ietf-pwe3-iccp-04
- IEEE 802.1Qbh：http://www.ieee802.org/1/pages/802.1bh.htm

作者介绍

蔡德忠（Dennis Cai），清华大学工程学士，北京大学计算机科学硕士，美国Depaul University数据通信硕士。2000年加入思科加州硅谷总部，现为思科核心技术产业部高级架构师，思科nV技术架构师。Dennis在运营商以太网领域申请了多项美国专利，合著了多篇IETF Draft。

王伯剑，毕业于天津南开大学计算机专业硕士，2005年加入思科中国，负责电信运营商网络与业务的顾问咨询，曾参与过多个国内国际大型网络项目的设计和实施，现为思科中国电信运营商事业部解决方案架构师。

北京
北京市朝阳区建国门外
大街2号北京银泰中心
银泰写字楼C座7-12层
邮编：100022
电话：(8610)85155000
传真：(8610)85155960

上海
上海市长宁区红宝石路500号
东银中心A栋21-25层
邮编：201103
电话：(8621)22014000
传真：(8621)22014999

广州
广州市天河区林和西路161号
中泰国际广场A塔34层
邮编：510620
电话：(8620)85193000
传真：(8620)85193008

成都
成都市滨江东路9号B座
香格里拉中心办公楼12层
邮编：610021
电话：(8628)86961000
传真：(8628)86961003

如需了解思科公司的更多信息，请浏览http://www.cisco.com.cn