



Mellanox ASAP²：加速交换和数据包处理

概述

对业务敏捷性、运营简单性和基础架构效率的追求催生了大量新技术，例如云、软件定义网络 (SDN)、网络功能虚拟化和云本机计算（面向现代分布式工作负载的容器封装、动态管理和微服务导向的计算环境）。这些技术的常见主题之一是它们都在挑战一项极限，即将多少个应用程序实例高效、安全和快速地带入特定的物理基础架构空间内。多个微服务、多个应用程序甚至多个租户共存需要高效的底层网络架构。通常，部署在服务器中的虚拟交换和路由软件是这种网络架构的关键要素之一，但因性能差和 CPU 开销高而面临挑战。

Mellanox 加速交换和数据包处理 (ASAP²) 解决方案将服务器/存储网络硬件的性能和效率与虚拟交换软件的灵活性相结合，为软件定义网络提供最高的总体基础架构效率、部署灵活性和操作简易性。

背景和挑战

虚拟交换因服务器虚拟化而生，因为 Hypervisor 需要能够实现虚拟机 (VM) 与外部世界之间的透明流量交换。最常用的虚拟交换软件解决方案之一是 Open vSwitch (OVS)，其目标是多服务器虚拟化部署。OVS 通常部署在云（例如 OpenStack）、SDN 和网络功能虚拟化 (NFV) 环境中，这些环境通常具有的特征是高度动态的端点以及高数据通信性能要求。这对于 NFV 尤其如此，其中应用程序本身是虚拟化网络功能 (VNF)，例如防火

墙、虚拟演进分组核心 (vEPC)、深度封包检测 (DPI) 等。

尽管虚拟交换机具有灵活性，但面临以下主要挑战：

- 糟糕的 I/O 性能
- 不可预测的应用程序性能
- 高 CPU 开销

糟糕的 I/O 性能

云构建者不再使用与 100Mb/s 或 1Gb/s 相同的旧服务器。根据 Crehan Research¹ 的研究，该行业即将达到拐点，其中 10Gb/s 以上速度的高速以太网总量将超过服务器级网卡（网络接口卡）总出货量的 50%。云服务提供商正在带头采用 25、40、50 甚至 100Gb/s 的服务器网卡来提高整体基础架构效率。对于高速服务器 I/O，每微秒可能有多个数据包到达，而普通 OVS 无法跟上其速度。在没有加速的情况下，OVS 在 10Gb/s 链路上可实现大约每秒 50 万个数据包 (pps) 的速度，而理论最大数据包速率可以达到 1,500 万 pps。在某些部署了电信 VNF 的实际应用场景中，流量主要由小型语音数据包构成，OVS 在 10Gb/s 接口上只能实现 1/80 的裸机 I/O 性能。

不可预测的应用程序行为

即使在 OVS 达到完全停止且无法转发更多数据包之前，运行速度也会大幅下降。队列累积，延迟飙升，并且可能会丢包。这反映在一些实时应用程序（如 VoIP）中，会以音质下降、卡顿或掉话等形式影响客户体验。应用程序性能的这种不确定性对云和服务提供商而言尤其成问题，因为他们无法向其客户提供强大的服务级别协议。

¹ 直至 2020 日历年的服务器级适配器和 LOM/控制器长期预测表，2016 年 1 月发布

高 CPU 开销

曾几何时，思科路由器中的所有数据包处理都是在所谓的慢速路径（也称为 CPU）中完成的。但是，现在没有任何著名网络供应商的路由器或交换机继续使用 CPU 进行数据包转发。相反，数据包处理和转发被卸载到硬件快速路径，通常在 ASIC 或网络处理器中实现。随着网络边缘被推进到服务器主机，裸机服务器可以实现高得多的数据包 I/O 性能，因为大多数数据包转发可以卸载到网卡。然而，对于计算和网络虚拟化，由于数据包需要在服务器中穿越路径，以及数据包格式发生变化，并非所有网卡都可以执行所需的卸载。当 CPU 执行数据包处理（包括校验和/CRC 计算和封装/解封装）时，多个 CPU 内核现在需要从应用程序处理转移到数据包处理。此数据包处理开销会导致正在运行其他工作负载的处理器紧急刹车。这导致应用程序性能显著下降，并降低了基础架构的整体效率。

解决方案 - ASAP²

从 ConnectX-4® 系列网卡开始，Mellanox 在服务器网卡硬件中通过 ASAP² 功能来支持加速虚拟交换功能。此独特功能是通过 ConnectX-4 硬件中的嵌入式交换机 (eSwitch) 实现的，该交换机可实现虚拟网卡（即 vNIC）之间的交换。利用网卡中内置的基于管线的可编程 eSwitch，ConnectX-4 可以在硬件中处理大部分数据包处理操作。这些操作包括 VXLAN 封装/解封、基于一组通用 L2-L4 报文头字段的数据包分类、QoS 和访问控制列表 (ACL)。ASAP² 基于这些增强的网卡硬件功能构建，提供可编程、高性能和高效的硬件转发层面，可与 SDN 控制层面无缝协作。它克服了与软件虚拟交换实现相关的性能和效率下降问题。在未来几代的 ConnectX 网卡（如 ConnectX-5 和 ConnectX-6）中，ASAP² 功能将得到进一步增强和拓展。有两种主要的 ASAP² 部署模式：ASAP² Direct 和 ASAP² Flex。

ASAP² Direct

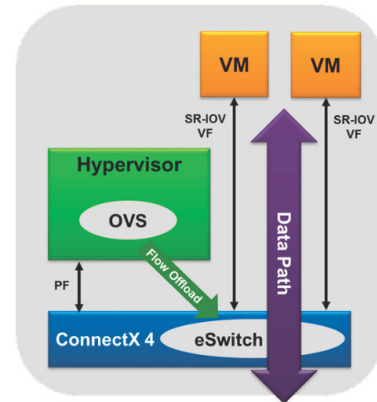
在此部署模式中，虚拟机通过单根 IO 虚拟化 (SR-IOV) 虚拟功能 (VF) 建立对 Mellanox ConnectX-4 网卡硬件的直接访问，从而在虚拟化环境中实现最高的网络 I/O 性能。

与传统 SR-IOV 实现相关的问题之一是它会完全绕过 Hypervisor 和虚拟交换机，并且虚拟交换机在 SR-IOV 模

式中不知道虚拟机的存在。因此，对于在服务器主机上使用 SR-IOV 的那些虚拟机，SDN 控制层面不能影响其转发层面。

ASAP² Direct 通过允许数据包处理操作从虚拟交换机卸载到 ConnectX-4 eSwitch 转发层面来克服此问题，同时保持 SDN 控制层面不受影响。对于 OVS，SDN 控

ASAP² Direct Full vSwitch offload

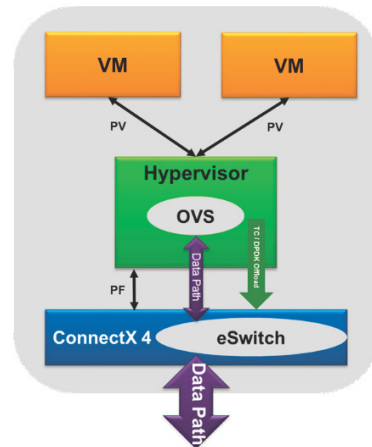


制层面保持不变：转发表和策略信息通过在用户空间中运行的 OVS vSwitchd 从相应的 SDN 控制器传递。然后根据用户定义的规则，转发表项可以编程写入 ConnectX-4 网卡的 eSwitch 而不是 OVS 内核模块。虚拟机 (VM) 实例通过 SR-IOV 连接到 ConnectX-4 网卡，并直接向/从网卡本身发送和接收数据包。

ASAP² Flex

在此部署模式中，虚拟机在半虚拟化模式下运行，并且仍然通过虚拟交换机满足其网络 I/O 需求。但是，

ASAP² Flex vSwitch acceleration



通过一组开放 API（例如 Linux 流量控制 (TC) 或数据路径开发工具包 (DPDK)），虚拟交换机可以将一些 CPU 密集型数据包处理操作卸载到 Mellanox ConnectX-4 网卡硬件，其中包括 VXLAN 封装/解封和数据包流量分类。ASAP² Flex 功能受具有所需固件的硬件和软件的支持，预计在 2017 年初面市。

ASAP² 解决方案优势

增强的云网络和 NFV 性能

ASAP² Direct 提供超越原始位吞吐量的优异小数据包性能。基准测试显示，在具有 25G 接口的服务器上，ASAP² Direct 加速的 OVS 实现了单流 3300 万数据包/秒 (MPPS) 的性能，并且 OVS 数据层面消耗的 CPU 计算资源接近于零；而在 ConnectX-4 Lx

eSwitch 中执行 VXLAN 封装/解封的 60,000 流实现了约 18 MPPS 的性能。这些性能结果是 DPDK 加速 OVS 的三到十倍。ASAP² 提供了以下两方面的最佳性能：软件定义的灵活网络可编程性，以及适用于从 10G 到 25/40/50/100G 的先进速度的高网络 I/O 性能。

总体基础架构效率

通过让网卡硬件承担 CPU 的 I/O 处理负荷，CPU 资源可以专门用于应用程序处理，从而提高系统效率。

部署灵活性

用于支持 ASAP² 的所有更改都将开放源码，并上传到 Linux、OVS 及其他社区。此外，ASAP² 对于应用程序是完全透明的——VNF 保持硬件独立性，不需要改变。这使得 ASAP² 部署简单易用且灵活。

进一步了解

了解 ASAP² 在云和 NFV 中战胜 DPDK 的三个方面

<http://www.mellanox.com/blog/2016/12/three-ways-asap2-beats-dpdk-for-cloud-and-nfv/>

观看 ASAP² Direct/OVS 卸载演示视频：

<https://www.youtube.com/watch?v=LuKuW5PAvwU>

了解有关 CloudX™ 优势的更多信息

<http://www.mellanox.com/solutions/cloud/benefits.php>



北京市朝阳区望京东园七区保利国际广场 T1 15 层
Tel: 010-5789 2000
www.mellanox.com