基于DPDK的Open vSwitch概述

原创：Robin G，姚磊 DPDK与SPDK开源社区 6月19日

作者简介

Robin Giller
英特尔网络平台集团的项目经理

本文是关于基于DPDK（数据面开发套件）的Open vSwitch[1]（OvS-DPDK）的概述，这是一种高性能，开源的虚拟交换机，本文还提供了有关OvS-DPDK功能的更为深入的技术文章的链接。本文是为希望深入了解OvS与DPDK集成的用户所撰写的。

注意：用户可以下载OVS主分支[2]或2.6分支[3]的压缩文件，及其安装步骤[4]。

（蓝色字部分链接见文末，下同）

PART 1 OvS-DPDK高级架构

OpenvSwitch是一个生产质量的多层虚拟交换机，在开源Apache*2.0 license下获得许可。它通过OpenFlow*协议及其OVSDB管理接口支持SDN控制语义。OpenvSwitch可以从openvswitch.org[5]，GitHub[6]*获得，也可通过Linux*发行版获得。

原有的 OpenvSwitch通常通过内核空间数据路径转发数据包（参见图1）。在内核数据路径中，交换机的“快速路径”包括一个对接收报文的转发和操作规则的简单流表。异常数据包（流中的第一个数据包）与内核快速路径表中的任何现有条目都不匹配，并被发送到用户空间守护程序进行处理（慢速路径）。在用户空间处理流中的第一个数据包之后，守护进程将更新内核空间中的流表，以便该流的后续数据包可以在快速路径中处理而不会发送到用户空间。根据此方法，对于大部分接收到的数据包，原始OvS可以消除在内核和

用户空间之间进行昂贵的上下文切换。但是，可达到的数据包吞吐量受到Linux网络协议栈的转发带宽限制，这不适用于需要高速率数据包处理的用例;例如，电信运营商场景。

DPDK是一组用户空间库，使用户能够创建高性能优化的数据包处理应用程序（可在DPDK.org[7]上找到相关信息）。实际上，它提供了一系列轮询模式驱动程序（PMD），可以绕过内核网络协议栈，在用户空间和物理接口之间直接传输数据包。通过消除内核网络协议栈的中断处理和遍历，可以显著提升转发性能。将OvS与DPDK集成后，交换机快速路径在用户空间中，异常路径与内核方案下的数据包路径相同。图1是DPDK与OvS的集成示意图。

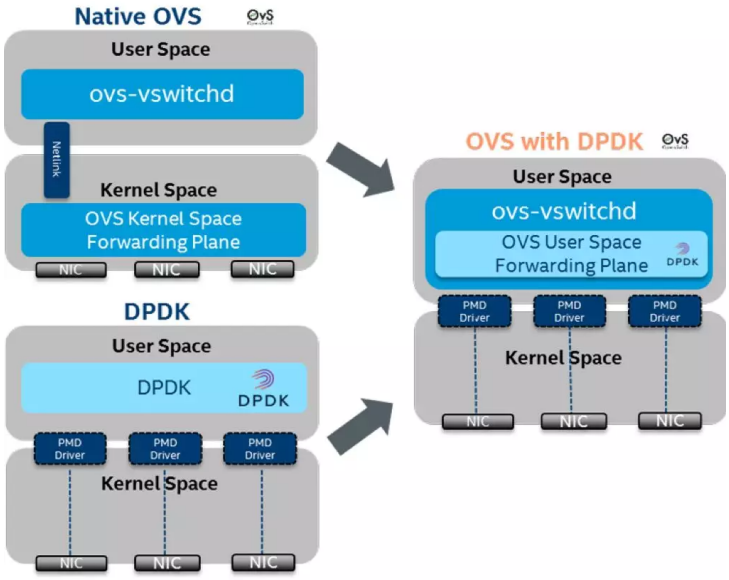

图1：DPDK（数据面开发套件）数据平面与原始Open vSwitch*的集成

下图2显示了OvS-DPDK的高级架构。OvS交换端口由网络设备（或netdevs）表示。Netdev-dpdk是一个DPDK加速过的网络设备，它通过三个独立的接口：一个物理接口（由DPDK中的librte_eth库处理）和两个虚拟接口（librte_vhost和librte_ring），使用DPDK加速交换机的I/O性能，这些物理和虚拟的接口会连接到虚拟交换机上。

其他OvS架构层提供了进一步的功能，并和例如SDN控制器相连接。Dpif-netdev提供用户空间转发，ofproto是实现OpenFlow交换的OvS库。它通过网络与OpenFlow控制器通信并通过ofproto提供软件和硬件接口的数据交换功能。ovsdb服务器维护该OvS实例的最新交换表信息并将其传送给SDN控制器。接下来的章节将提供交换/转发表的详细信息，更多有关OvS体系结构的信息[8]，可通过openvswitch.org网站获取。

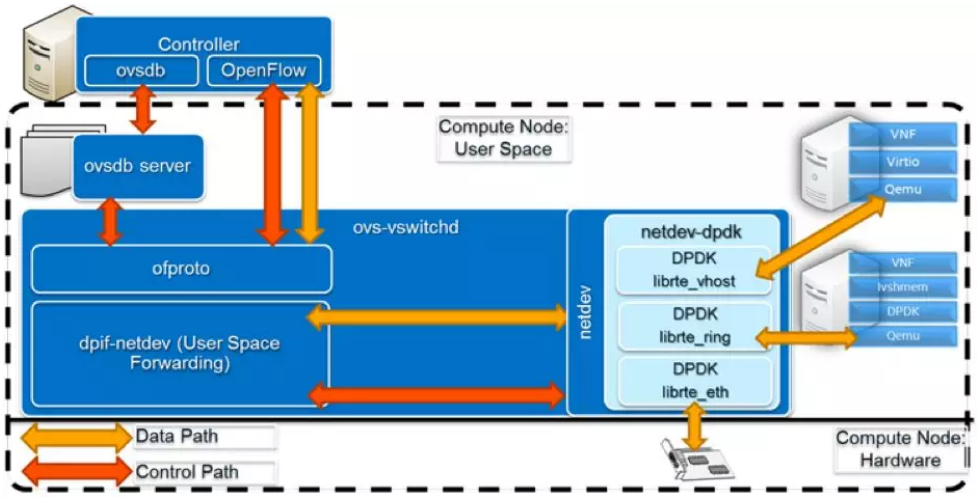

图2：使用DPDK（数据面开发套件）高级架构的OpenvSwitch *

PART 2 OvS-DPDK交换表层次结构

从物理或虚拟接口进入OvS-DPDK的数据包根据其头部字段来确定唯一标识符或散列，然后将其与三个主要交换表之一中的条目进行匹配，三个主要交换表是：完全匹配缓存（EMC），数据路径分类器（dpcls）或ofproto分类器。除非找到匹配，否则数据包的标识符将按顺序遍历这三个表，若找到匹配，则将执行表中匹配规则指示的相应操作并且在完成所有动作后将数据包转出交换机。该方案如图3所示。

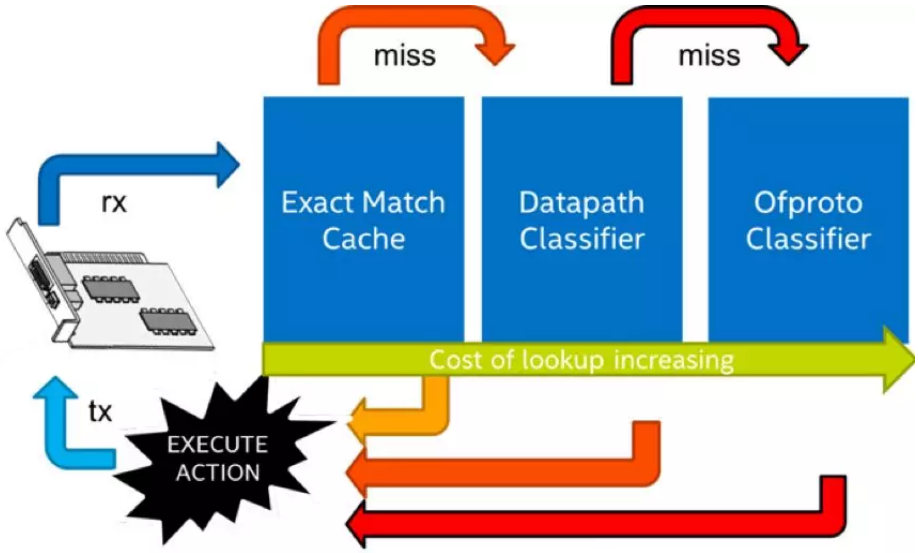

图3：使用DPDK（数据面开发套件）交换表层次结构的OpenvSwitch *

这三个表具有不同的特性和相应吞吐量性能/延迟。EMC为有限数量的表条目提供最快的处理。数据包的标识符必须与此表中的所有字段的条目完全匹配 ，包括源IP和端口的5元组，目标IP和端口，以及协议，这是为了最高速度处理，否则它将未命中EMC并传递给dpcls。dpcls包含更多表条目（排列在多个子表中）并启用数据包标识符的通配符匹配（例如，指定目标IP和端口，但允许任何源）。这提供了大约一半的EMC吞吐量性能，并

且可以满足更多的表条目。将dpcls中匹配的数据包流配置在EMC中，可以以最高速度处理具有相同标识符的后续分组。

未命中 dpcls会导致数据包被发送到ofproto分类器，以便OpenFlow控制器可以决定数据包操作。此路径的性能最低，比EMC慢10倍以上。ofproto分类器中的匹配会导致新表条目建立在切换更快的表中，以便可以更快地处理同一流中的后续数据包。

▼

PART 3 OvS-DPDK功能和性能

在撰写本文时，OvS主代码分支上提供了以下高级的OvS-DPDK功能：

- v16.07的DPDK支持（每个新DPDK版本支持的版本增量）
- vHost-user支持
- vHost重连
- vHost多队列
- 本地隧道支持：VxLAN，GRE，Geneve
- VLAN支持
- MPLS支持
- Ingress/egress/QoS策略
- 巨帧支持
- 连接跟踪
- 统计：DPDKvHost和扩展的DPDK统计数据
- 调试：DPDKpdump支持
- 链接聚合
- 链接状态
- VFIO支持
- DPDK端口的ODL/OpenStack检测
- vHost-user NUMA感知

图4突出显示了原有OvS和OvS-DPDK之间的最新性能比较。这显示了Phy-OvS-Phy使用场景下的每秒数据包吞吐量，表明OvS-DPDK比原有的OvS性能提升约为10倍，在启用英特尔®超线程技术（英特尔®HT技术）的情况下增加到约12倍（标记为1C2T，或图中带有两个逻辑线程的物理核心）。类似地，Phy-OvS-VM-OvS-Phy用户场景表明OvS-DPDK相对于原有OvS的性能提高了约9倍。

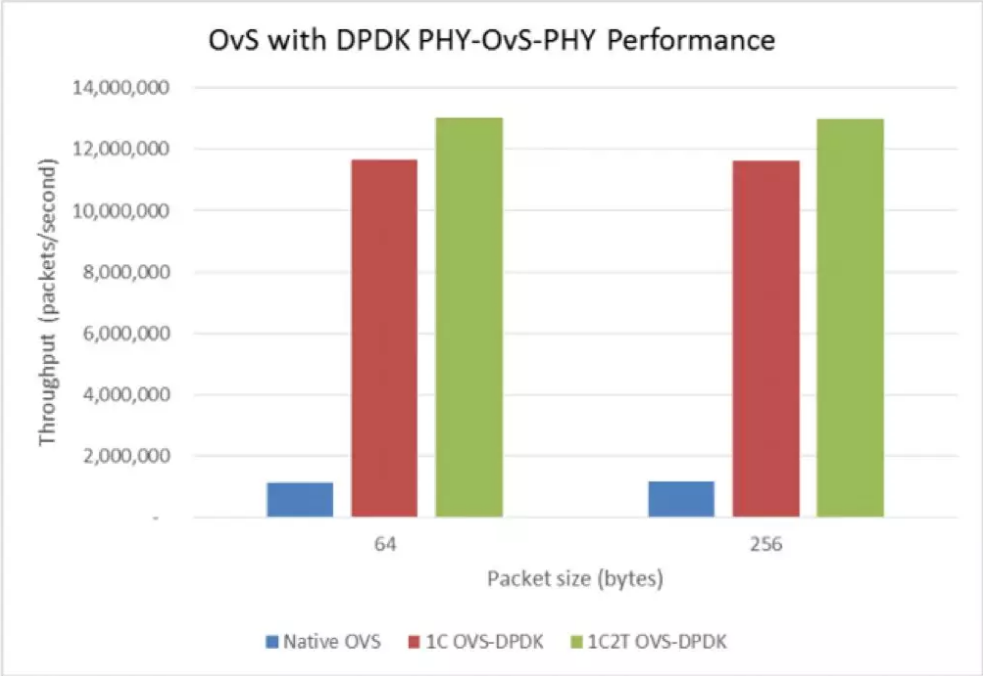

图4：性能比较 – 原有OpenvSwitch *（OvS）和带数据平面开发套件的OvS

此数据的硬件和软件配置以及更多用例结果可在英特尔®开放式网络平台（英特尔®ONP）性能报告[9]中找到。

OvS-DPDK获取方式

OvS-DPDK可以在openvswitch.org[10]存储库中找到，也可以通过如下所示Linux发行版获得。最新的里程碑版本是OvS2.6（2016年9月），发布时间为6个月。

代码可供下载，如下所示：OvS主分支[11];OvS 2.6版本分支[12]。主分支的安装步骤[13]以及2.6版本分支的安装步骤[14]都可获得。

基于DPDK的OvS打包版本可从以下网址获得：

红帽*OpenStack平台[15]

Ubuntu*[16]

Mirantis * OpenStack[17]

Open Platform for NFV*[18]

附加信息

要了解有关OvS-DPDK的更多信息，请查看以下视频和文章，来自英特尔®开发人员专区，01.org，英特尔®网络构建商和英特尔®网络建设者大学。

用户指南

使用基于DPDK的OvS进行VM间NFV应用[19]

在Ubuntu上使用基于DPDK的OvS[20]

开发者指南

基于DPDK的OpenvSwitch - 如何使用DPDK数据路径构建和安装Open vSwitch[21]

使用基于DPDK的OpenvSwitch - 包括高级性能调优信息[22]

文章和视频

基于DPDK的OvS的速率限制配置和使用[23]

基于DPDK的OvS的QoS配置和使用[24]

基于DPDK的OvS中vHost user多队列配置[25]

OvS-DPDK中vhost user的NUMA感知[26]

基于DPDK的OvS中的Pdump[27]

在OpenStack中使能OvS-DPDK[28]

基于DPDK的Open vSwitch *中的巨帧[29]

基于DPDK的vSwith中vhost-user客户端模式[30]

OVS-DPDK数据通分类器- 第1部分[31]

OvS-DPDK数据通路分类器- 第2部分[32]

OvS-DPDK中的链路聚合配置和使用[33]

使用英特尔®VTuneAmplifier分析OvS-DPDK中的性能瓶颈[34]

在DevStack中使基于OvS和DPDK的Neutron[35]

使用OVS-DPDK构建和测试简单的NFV 跨VM使用场景（YouTube视频系列）[36]

OvS与DPDK里程碑发布网络研讨会

在OpenStack中使用DPDK启用OvS2.5.0[37]

在OpenStack中使用DPDK启用OvS2.4.0[38]

在OpenStack中使用DPDK启用OvS2.6.0[39]

INB大学

OvS-DPDK深入研究[40]

DPDK Open vSwitch：加速虚机访问路径[41]

白皮书

OvS-DPDK使能SDN和NFV[42]

有问题？请随时关注Open vSwitch讨论邮件[43]线程上的问答。

【序号链接网址】

1. https://software.intel.com/en-us/user/1386481
2. http://docs.openvswitch.org/en/latest/intro/install/dpdk/
3. https://github.com/openvswitch/ovs/archive/master.zip
4. http://github.com/openvswitch/ovs/archive/branch-2.6.zip
5. https://github.com/openvswitch/ovs/blob/branch-2.6/INSTALL.DPDK.md
6. http://openvswitch.org/
7. https://github.com/openvswitch/ovs/blob/master/Documentation/topics/porting.rst
8. http://dpdk.org/
9. https://github.com/openvswitch/ovs/blob/master/Documentation/topics/porting.rst
10. https://download.01.org/packet-processing/ONPS2.1/Intel_ONP_Release_2.1_Performance_Test_Report_Rev1.0.pdf
11. http://openvswitch.org/
12. https://codeload.github.com/openvswitch/ovs/zip/master
13. https://codeload.github.com/openvswitch/ovs/zip/branch-2.6
14. https://github.com/openvswitch/ovs/blob/master/Documentation/intro/install/dpdk.rst
15. https://github.com/openvswitch/ovs/blob/branch-2.6/INSTALL.DPDK.md
16. https://access.redhat.com/products/red-hat-openstack-platform"\t "_blank
17. http://releases.ubuntu.com/16.04/" \t"_blank
18. https://www.mirantis.com/products/mirantis-openstack-software/"\t "_blank
19. https://www.opnfv.org/brahmaputra"\t "_blank
20. https://software.intel.com/en-us/articles/using-open-vswitch-with-dpdk-for-inter-vm-nfv-applications
21. https://software.intel.com/en-us/articles/using-open-vswitch-with-dpdk-on-ubuntu
22. http://docs.openvswitch.org/en/latest/intro/install/dpdk/
23. http://docs.openvswitch.org/en/latest/howto/dpdk/
24. https://software.intel.com/en-us/articles/rate-limiting-configuration-and-usage-for-open-vswitch-with-dpdk
25. https://software.intel.com/en-us/articles/qos-configuration-and-usage-for-open-vswitch-with-dpdk
26. https://software.intel.com/en-us/articles/configure-vhost-user-multiqueue-for-ovs-with-dpdk
27. https://software.intel.com/en-us/articles/vhost-user-numa-awareness-in-open-vswitch-with-dpdk
28. https://software.intel.com/en-us/articles/dpdk-pdump-in-open-vswitch-with-dpdk
29. https://01.org/openstack/blogs/stephenfin/2016/enabling-ovs-dpdk-openstack
30. https://software.intel.com/en-us/articles/jumbo-frames-in-open-vswitch-with-dpdk
31. https://software.intel.com/en-us/articles/vhost-user-client-mode-in-open-vswitch-with-dpdk
32. https://software.intel.com/en-us/articles/ovs-dpdk-datapath-classifier
33. http://software.intel.com/en-us/articles/ovs-dpdk-datapath-classifier-part-2
34. http://software.intel.com/en-us/articles/link-aggregation-configuration-and-usage-in-open-vswitch-with-dpdk

35. https://software.intel.com/en-us/articles/analyzing-open-vswitch-with-dpdk-bottlenecks-using-vtune-amplifier
36. https://software.intel.com/en-us/articles/using-open-vswitch-and-dpdk-with-neutron-in-devstack
37. https://www.youtube.com/playlist?list=PLg-UKERBljNwc7kz6bsRnwPb4kTBSldLs
38. https://www.brighttalk.com/webcast/12229/209935"\t "_blank
39. https://www.brighttalk.com/webcast/12229/194949"\t "_blank
40. http://www.brighttalk.com/webcast/12229/232617/open-vswitch-with-dpdk-in-ovs-2-6-0
41. https://networkbuilders.intel.com/university/course/open-vswitch-with-dpdk-architectural-deep-dive
42. https://networkbuilders.intel.com/university/course/dpdk-open-vswitch-accelerating-the-path-to-the-guest"\t "_blank
43. https://networkbuilders.intel.com/docs/open-vswitch-enables-sdn-and-nfv-transformation-paper.pdf
44. http://mail.openvswitch.org/mailman/listinfo/discuss

转载须知

DPDK与SPDK开源社区公众号文章转载声明

推荐阅读

一文读懂SPDK用户态hotplug处理
缓存助力存储加速–OCF与SPDK介绍及用法
TestPMD使用中的 ixgbe filter 设置
DPDK发布19.05
SPDK发布19.04

DPDK
SPDK
DPDK与SPDK开源社区

长按二维码关注 获取最新资讯

阅读原文