

# 基于virtio-user的新exception path方案

原创 DPDK开源社区 2017-03-08

作者 许茜, 谈鉴锋



↑ 点击蓝字，轻松关注

在DPDK使用环境中，物理网卡收到的包都绕开内核，直接到达DPDK应用中。但是，有些时候，用户希望把某些包（如控制报文）放到内核网络协议栈进行处理，这个路径在DPDK中被称作exception path。

## 现有方案

现有的exception path方案主要有三个：

1. KNI，作为目前DPDK用户使用的主要方案，其通过内核模块构造了一个虚拟网络接口，并且通过FIFO队列和用户态的DPDK应用交换数据包。该方案的缺点是KNI内核模块无法upstream，维护代价较大。详情请参见：

[http://dpdk.org/doc/guides/prog\\_guide/kernel\\_nic\\_interface.html](http://dpdk.org/doc/guides/prog_guide/kernel_nic_interface.html)。

2. Tun/tap或者pcap PMD，这种方式的主要问题是这些端口的PMD需要调用syscall进入到内核态来收发包，这样的切换会导致DPDK应用性能大幅下降。

关于tun/tap PMD，详情请参见：

<http://dpdk.org/doc/guides/nics/tap.html>；

关于pcap PMD，详情请参见：

[http://dpdk.org/doc/guides/nics/pcap\\_ring.html](http://dpdk.org/doc/guides/nics/pcap_ring.html)。

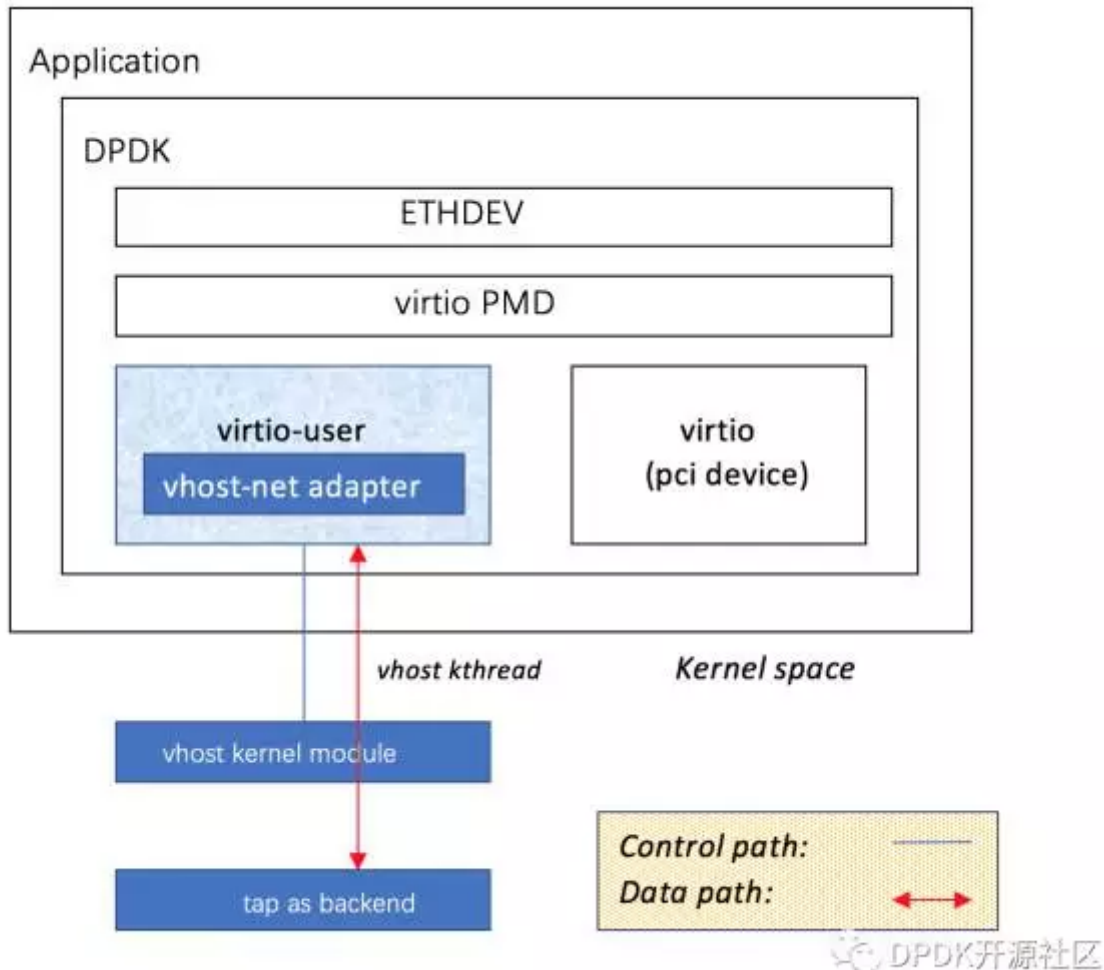
3. Flow Bifurcation，该方式利用物理网卡SR-IOV能力虚拟出多张网卡（一个PF和多个VF），内核和DPDK应用各拿着一些，并且利用网卡的分流引擎将特定网络流导向内核驱动所控制的虚拟网卡。这种方式依赖于硬件的包分类和包过滤能力，不够灵活。

详情请参见：

<https://dpdksummit.com/Archive/pdf/2016Userspace/Day02-Session05-JingjingWu-Userspace2016.pdf>。

## 一种新方案

如大家所知，virtio作为一种半虚拟化方案服务于虚拟机网络、存储等。在DPDK v17.02，我们推出了virtio-user这样一种虚拟设备，结合vhost-kernel后端驱动（upstream到Linux的内核模块）来为DPDK应用提供一种快速、灵活的exception path方案。其设计如下图所示：



Virtio-user是virtio PMD的虚拟设备，启动DPDK virtio-user，系统就会创建一个内核态的虚拟设备tap，tap通过vhost-kthread和virtio-user进行数据的发送接收，这是数据面；vhost-net的内核模块是virtio-user的控制面，发送接收一些控制消息。这样一来，从DPDK收到的包进入到virtio-user，通过vhost-kthread进入到tap设备，tap设备支持内核协议栈，很好地实现了exception path的包处理。

### 亮点分析

从维护角度来看，本方案所依赖的vhost.ko和vhost-net.ko都是早已upstream的内核模块，不需要维护out-of-tree的内核模块；从灵活性来看，本方案不依赖任何硬件功能；从线程模型来看，和KNI相似，本方案只需将数据包放到virtio ring上，数据拷贝操作由vhost kthread来完成。从网络功能来看，vhost-net本来就是为网络而生的，能通过checksum计算和验证、数据包分片offload到物理网卡来进行。由于对Multi-seg数据包的支持，和KNI的方案相比，本方案将iperf的性能提升了2倍以上。

这么来看，virtio-user+ vhost-net作为一种新的exception path方案还是有很强的优势。测试步骤请参见：

[http://dpdk.org/doc/guides/howto/virtio\\_user\\_as\\_exceptional\\_path.html](http://dpdk.org/doc/guides/howto/virtio_user_as_exceptional_path.html)

还不赶快来试试！

### 作者简介

**许茜**，浙江大学信电系硕士，现为英特尔网络处理事业部软件测试工程师，主要负责DPDK 相关的虚拟化测试和性能测试。

**谈鉴锋**，清华大学计算机硕士，现为英特尔工程师，主要从事于DPDK中半虚拟化网卡virtio的驱动开发，目前专注于使用DPDK加速容器网络以及推动容器的NFV技术。



DPDK开源社区



长按指纹识别二维码关注

[阅读原文](#)

[投诉](#)