



FTF 2016
TECHNOLOGY FORUM CHINA

KVM虚拟化：QorIQ平台上利用I/O虚拟化实现VNFS

FTF-NET-N1844

BHARAT BHUSHAN
首席主管工程师
DIANA CRĂCIUN
软件工程师
XIN-XIN YANG
软件研发总监

2016年9月

公开使用



软件产品和服务

开发工具

- CodeWarrior

运行时态产品

- VortiQa软件解决方案

CodeWarrior
QorIQ

VortiQa



集成服务

- 安全咨询
- 强化Linux

解决方案参考

- 物联网网关
- OpenWRT+

Linux®服务

- 商业支持

- 性能调整



加快客户产品上市时间



交付商用软件、支持、服务和解决方案



简化与恩智浦的软件合作



创造成功!

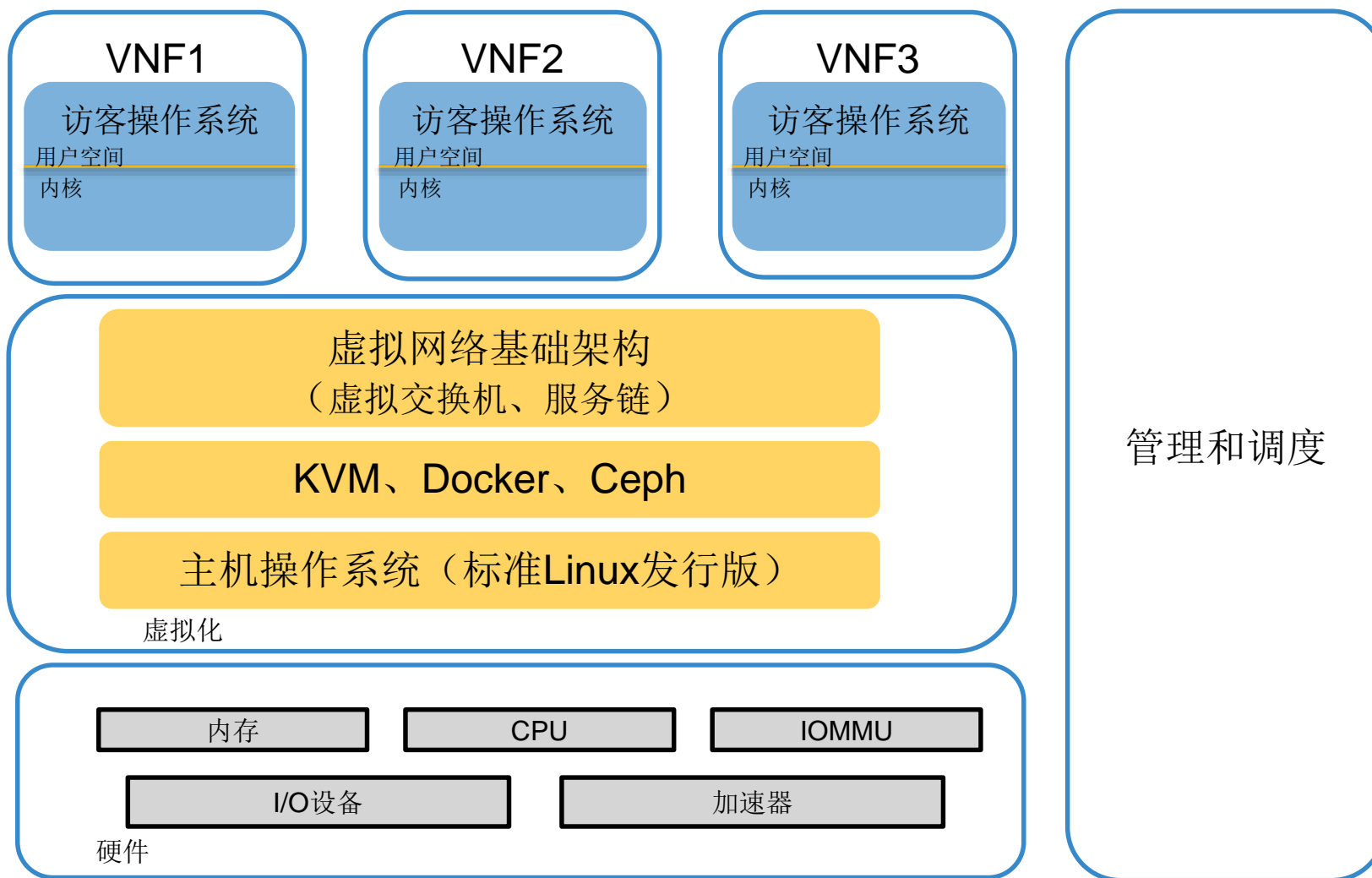


议程

- 虚拟化概述
- I/O虚拟化
- 直接分配
- VirtIO
- 结论

虚拟化概述

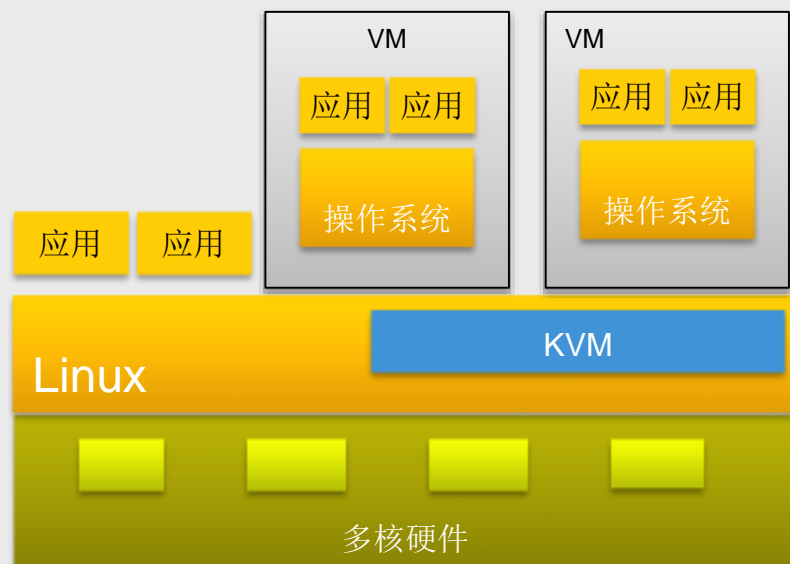
NFV和VNF



恩智浦虚拟化解决方案

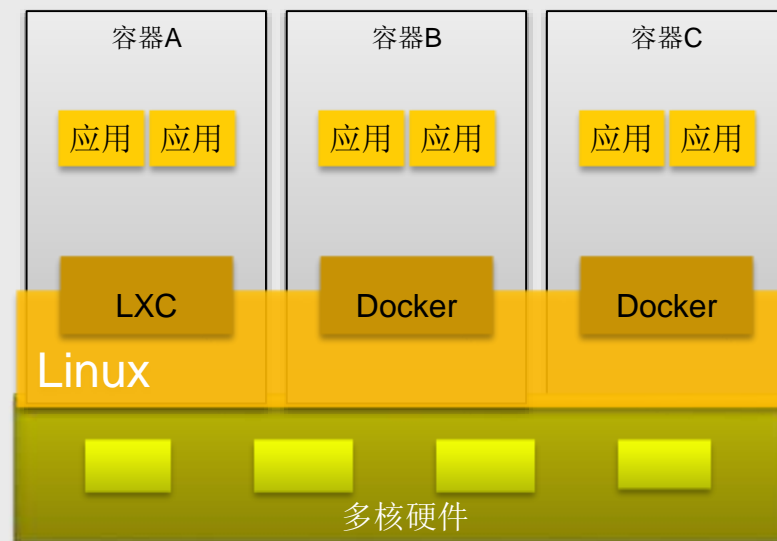
KVM

- Linux ® Hypervisor
- 资源虚拟化/超额申请
- 开源
- 采用Qemu用户空间仿真

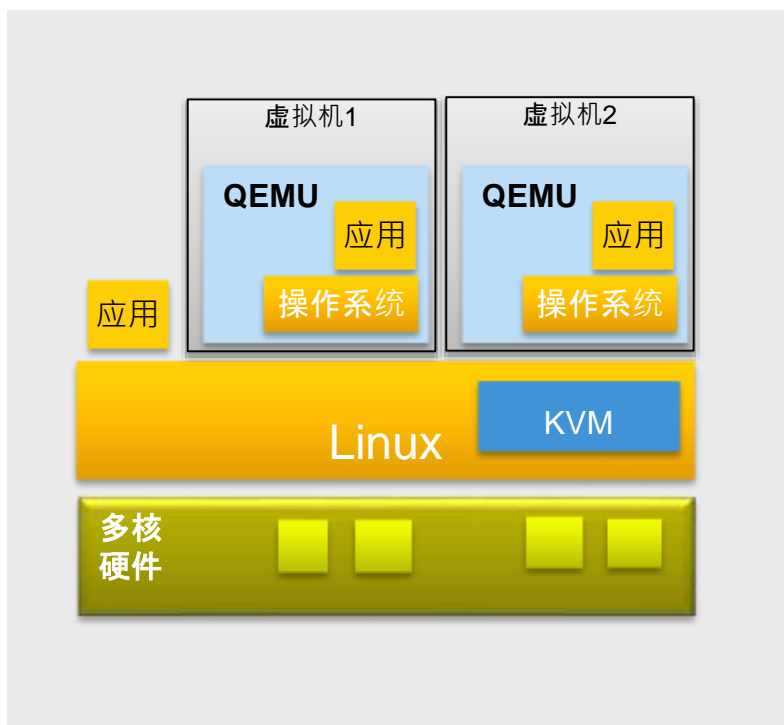


操作系统虚拟化

- 轻量级开销
- Linux ®中的隔离和资源控制
- 减小隔离性（内核共享）

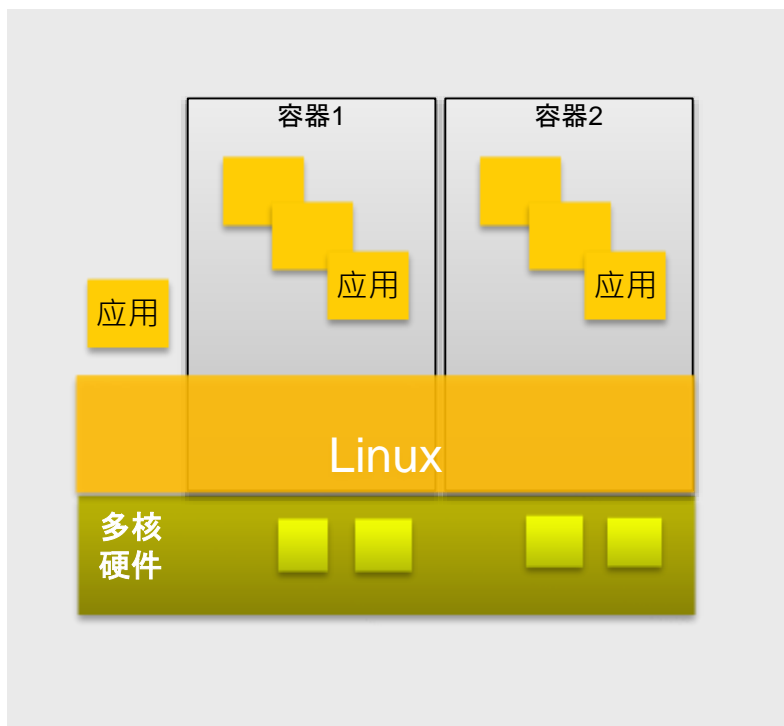


KVM/QEMU



- KVM/QEMU – 基于Linux®内核的开源虚拟化技术
- KVM是一个Linux内核模块
- QEMU是一个使用KVM进行加速的用户空间仿真器
- 同时运行虚拟机和Linux应用程序
- 操作系统无需更改或仅需小幅更改
- 虚拟I/O功能
- 直接/直通I/O – 为VM分配I/O设备

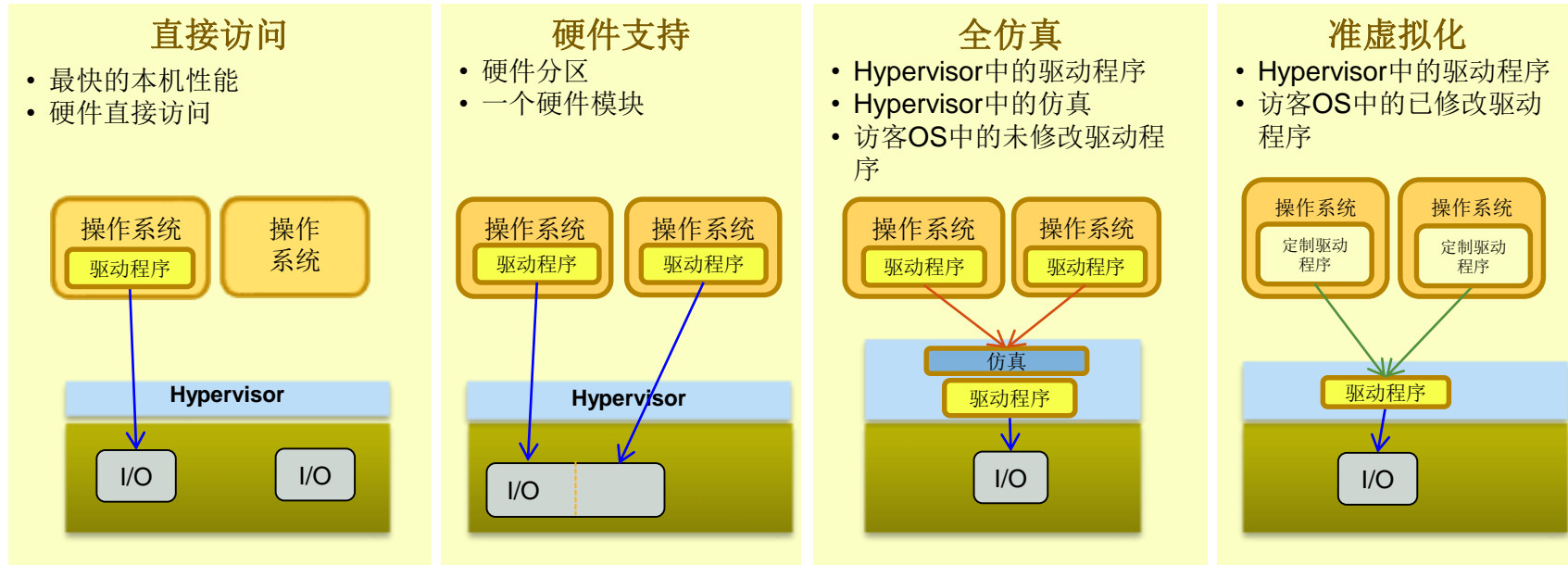
Linux容器



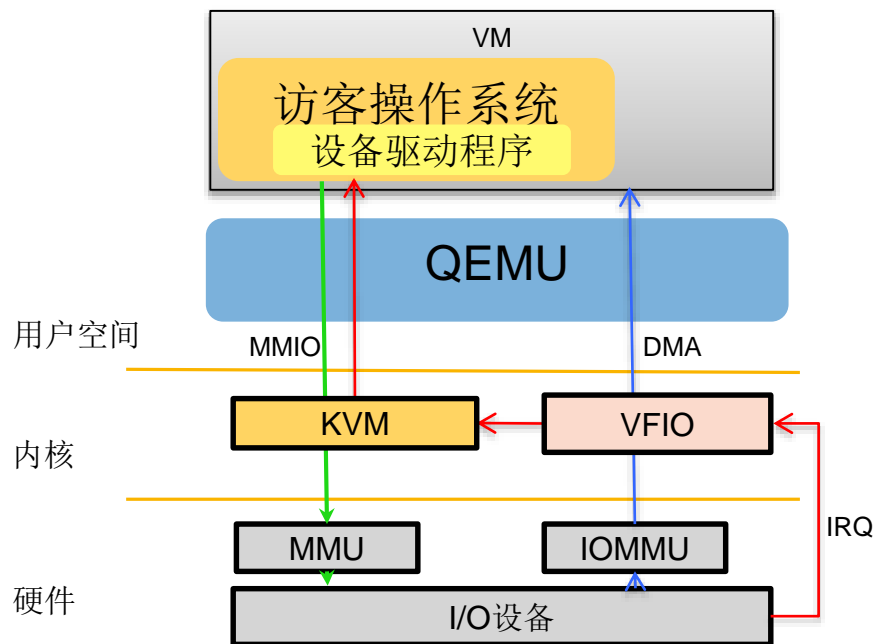
- 操作系统级虚拟化/进程级虚拟化
- 主机和访客单一内核，虚拟化用户空间 – 操作系统隔离
- 以低开销、轻量级和安全的方式将Linux应用程序划分在不同的域中
- 根据每个域进行资源利用控制 – CPU、内存、I/O带宽
- 多资源情况 – 命名空间
 - 进程 – 进程树
 - 网络 – 网络堆栈（netdev、socket families、FDB）
- 基于包括内核组件（cgroup、命名空间）在内的技术和用户空间工具（LXC、libvirt、Docker）的集合

I/O虚拟化

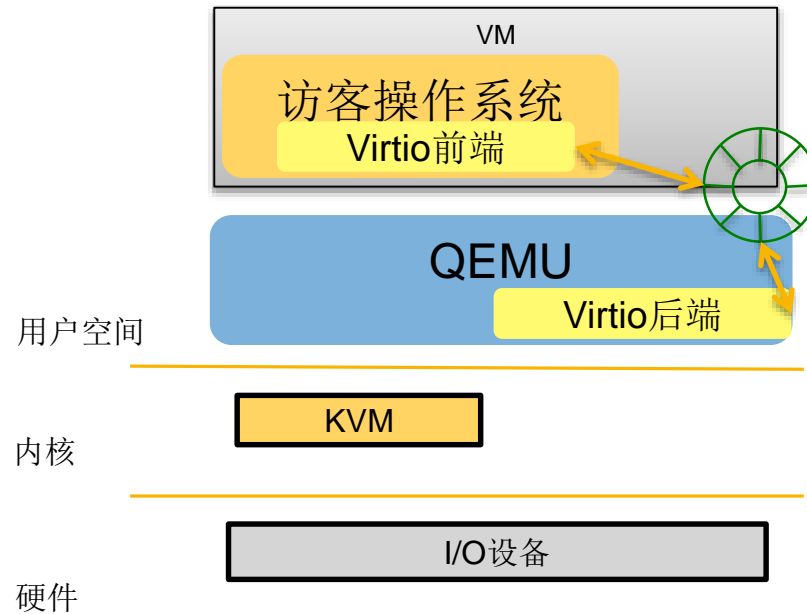
虚拟环境中的设备使用



KVM/Linux中的设备使用

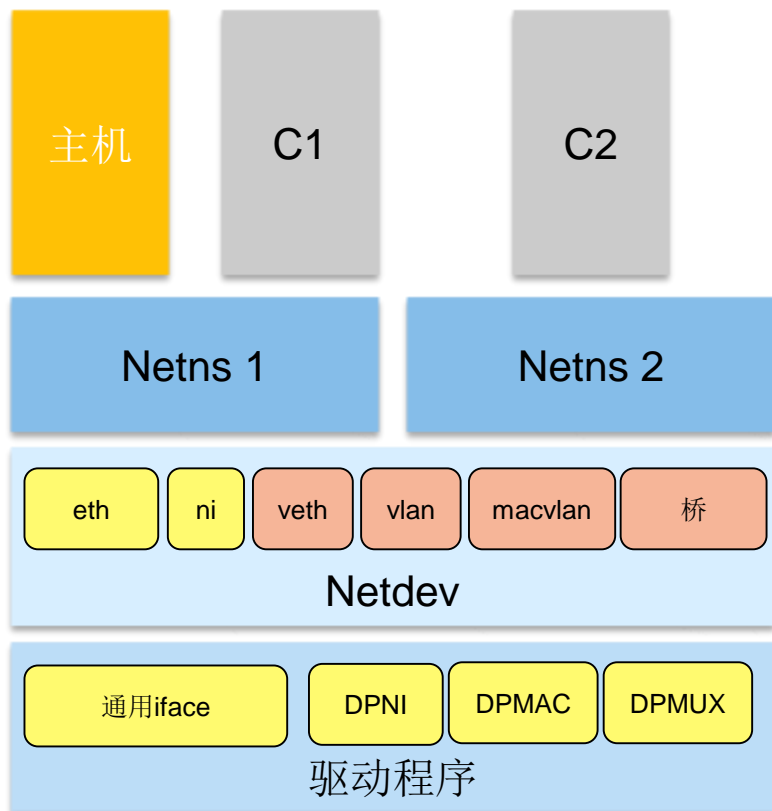


VFIO (简化视图)



Virtio (简化视图)

容器中的设备使用

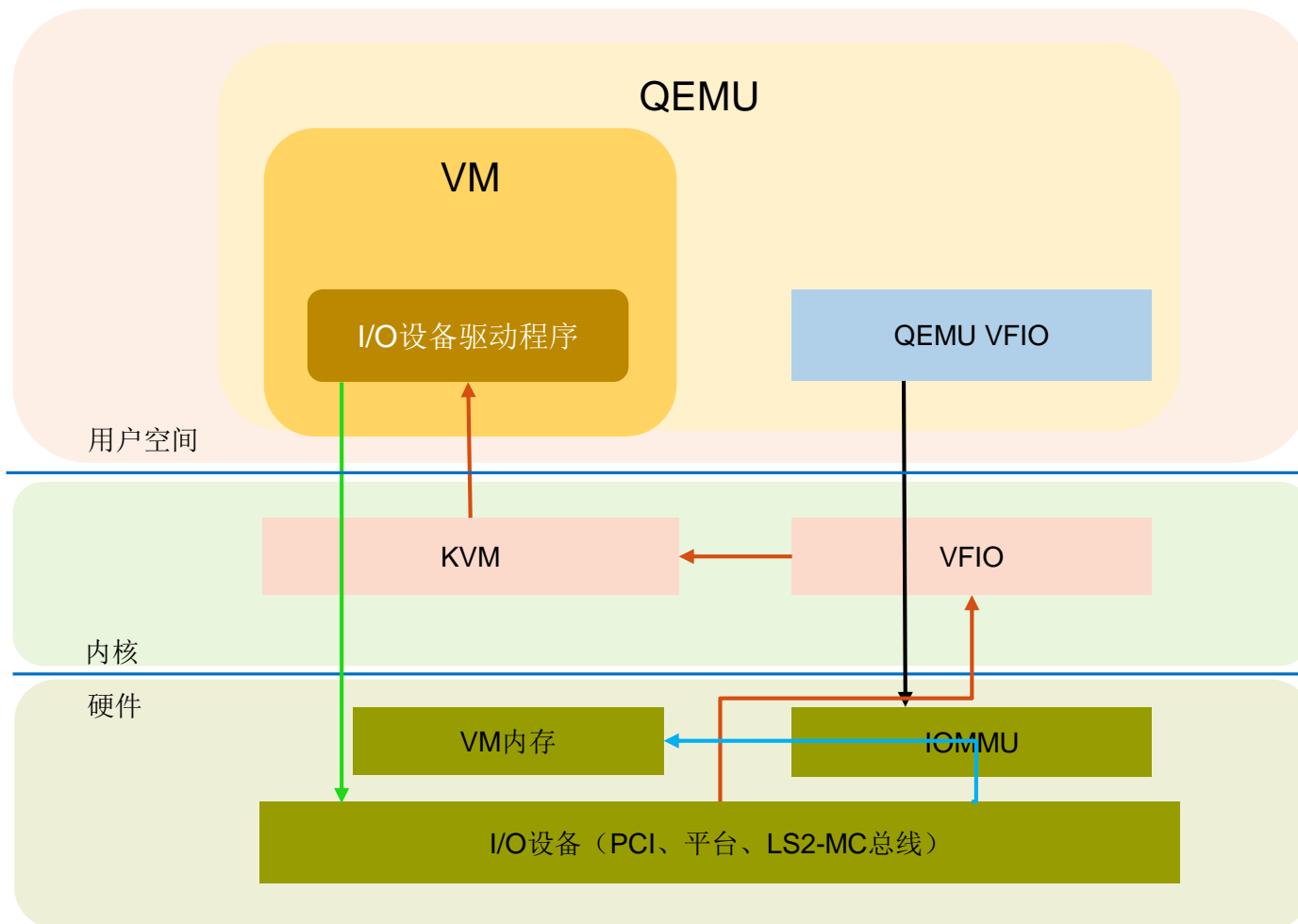


- 每个容器（用户空间实例）具有一个网络命名空间
 - 多个容器可共享相同网络命名空间
- 每个netdev属于一个网络命名空间
- netdev可以是：
 - 物理：具有关联的硬件设备或抽象
 - 虚拟：完全是软件（veth、vlan、桥等）
- 虚拟netdev开销较低 – 差异源于技术细节
 - 桥：内核切换
 - MACVLAN：MAC级VLAN
 - VETH：IP级软件对
- 混合和匹配

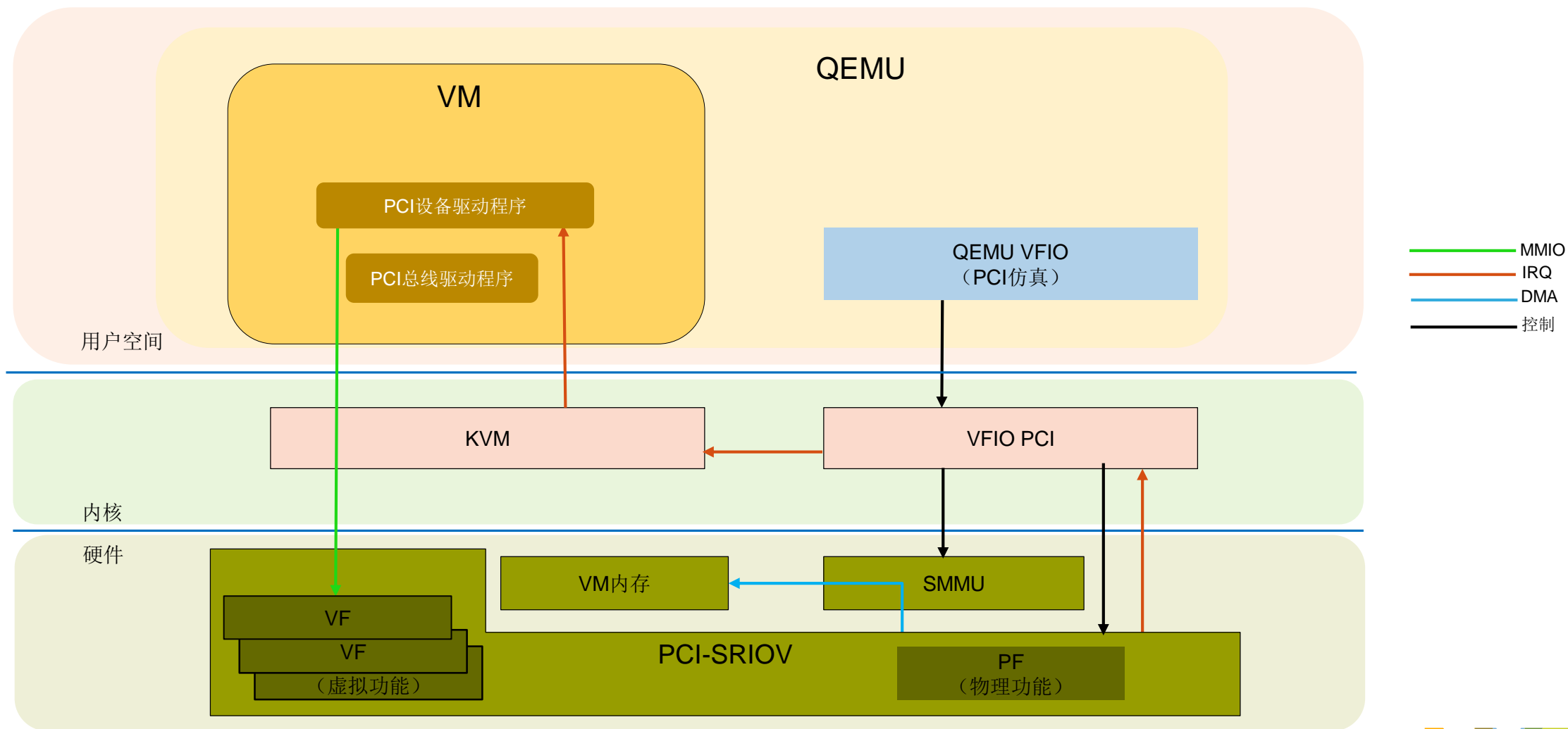
直接分配

VFIO简介

- VFIO（虚拟功能IO）
 - Linux用户空间驱动程序基础架构
 - 强化IOMMU保护
- VFIO提供
 - 设备使用权限（mmap()设备MMIO区域）
 - IOMMU编程接口
 - 高性能中断支持
- 总线支持
 - PCI、平台设备、LS2 MC总线

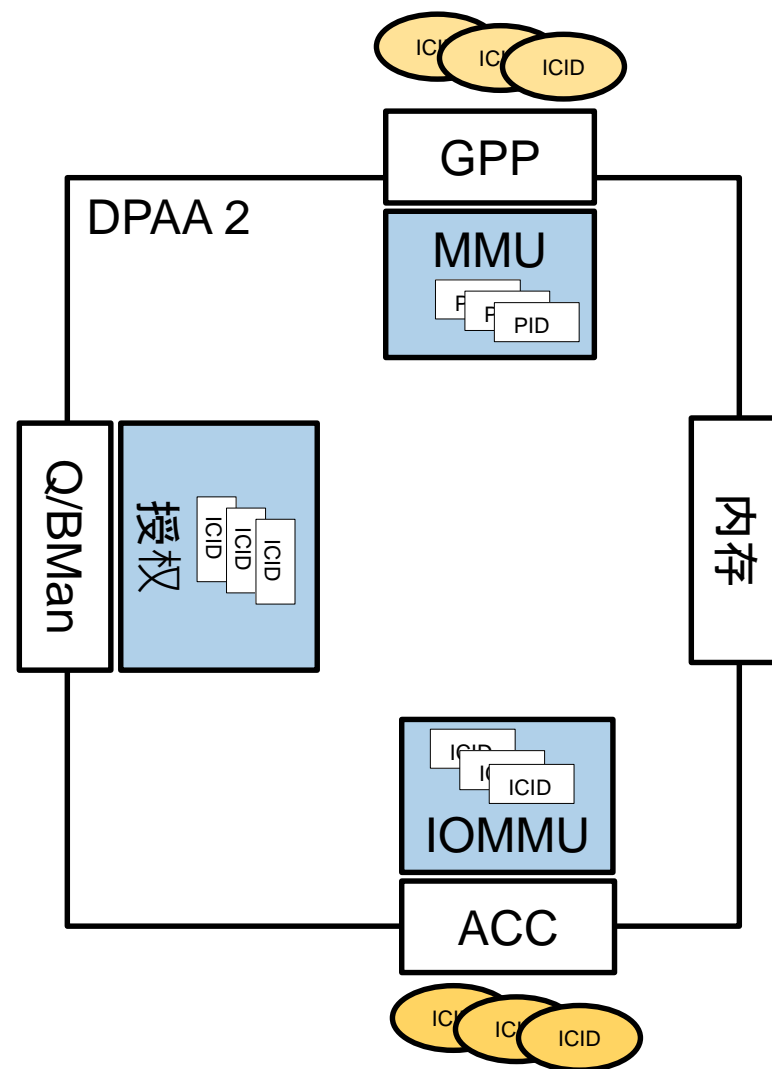


PCI设备直接分配到VM

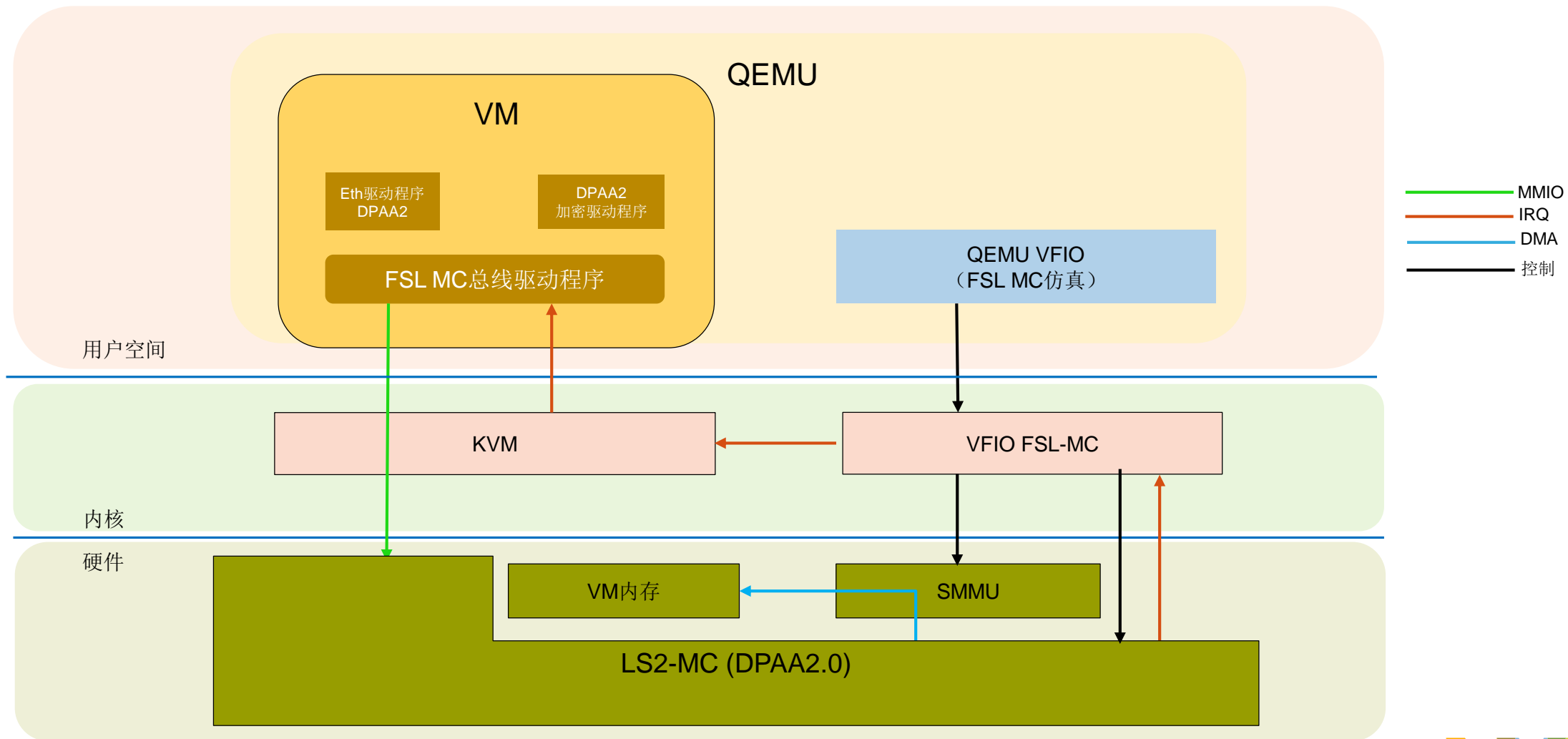


DPAA2可实现安全的直接分配

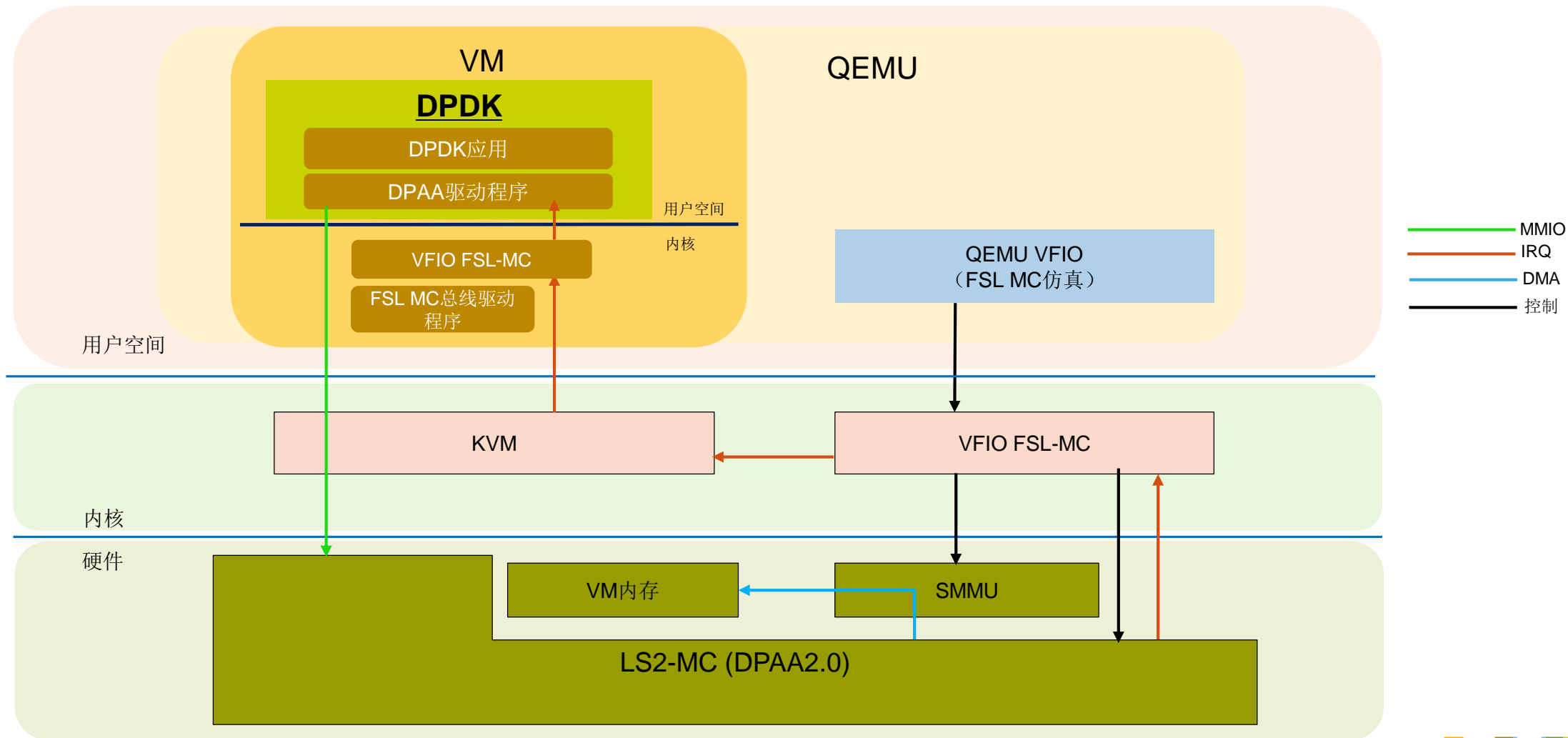
- Management Complex(MC)针对通过MC将资源分配至各种软件环境进行了优化
 - Linux MC总线
 - 资源管理工具
- 用户空间的IOMMU转换和保护 (DPDK和QEMU)
 - ICID (StreamID)
 - MC总线与VFIO集成
 - 设备复位
- DPAA通过授权表确保安全



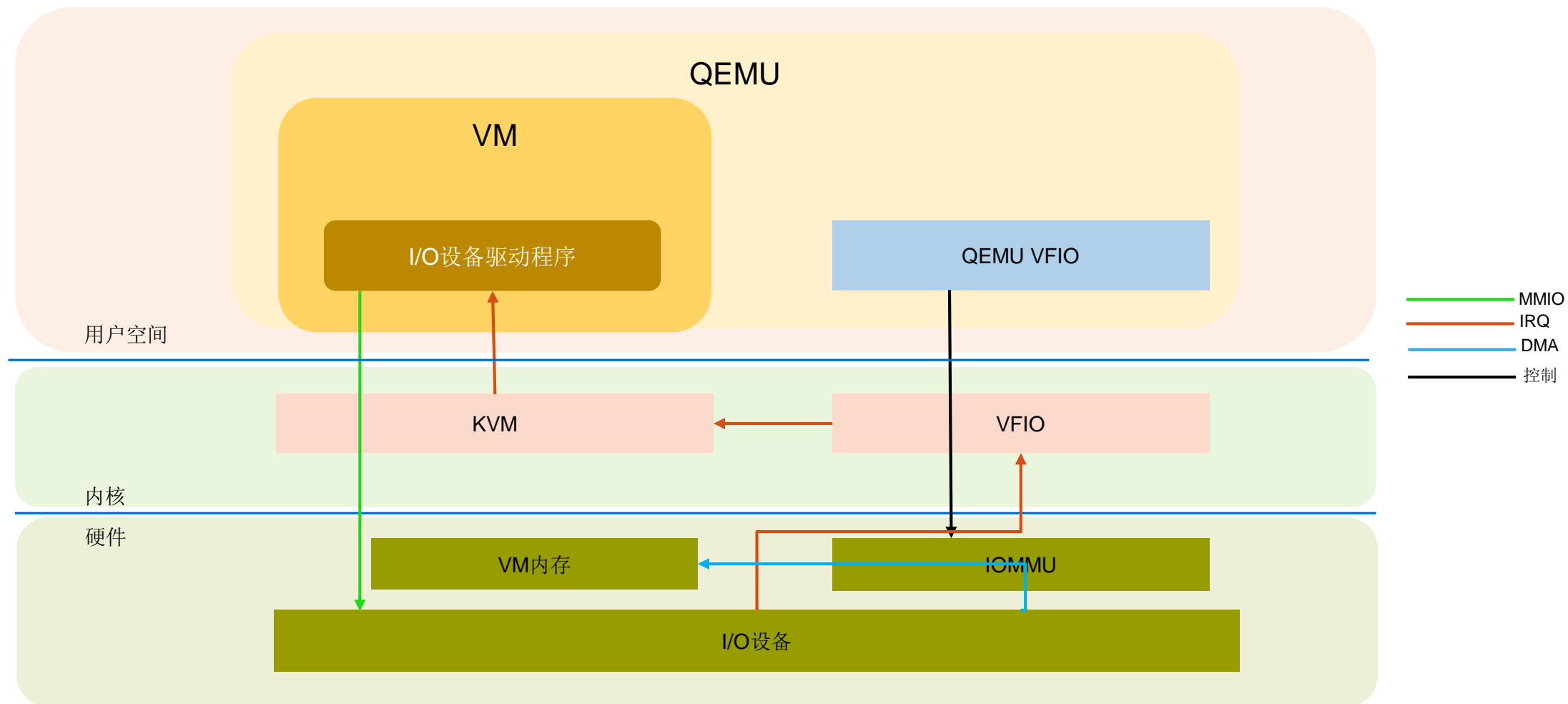
DPAA2设备直接分配到VM



DPAA2设备直通至VM中的DPDK



平台设备直接分配



VIRTIO详情

虚拟I/O设备

Virtio系列设备

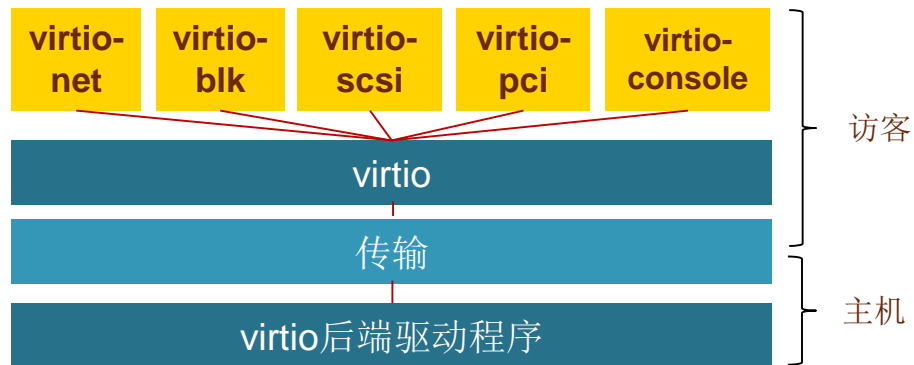
- 在虚拟环境中找到
- 其设计看起来像物理设备
- 使用访客操作系统标准驱动程序和发现机制
- 由OASIS技术委员会定义的规格

Virtio规格用途

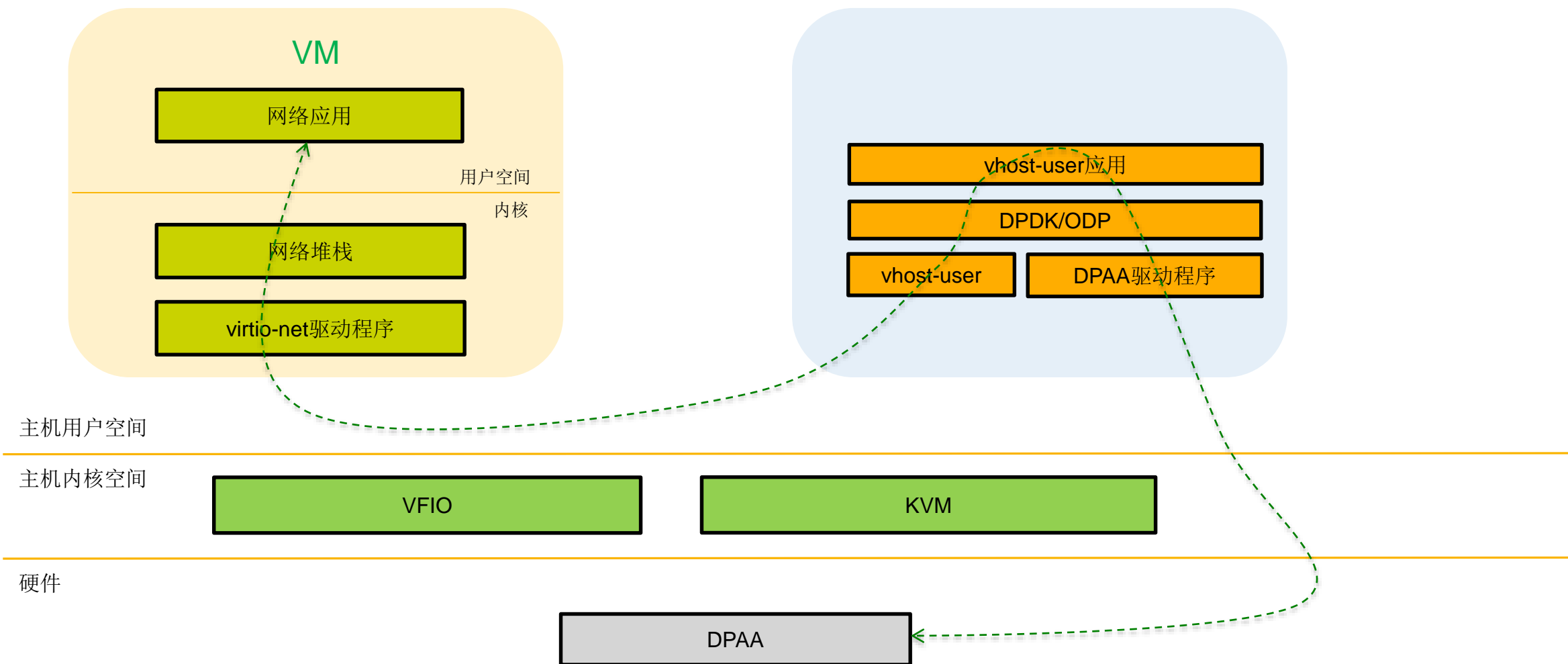
- 简单 - 采用中断和DMA的常规总线机制
- 高效 - 使用输入和输出的描述符环，以避免缓存效应
- 标准 - 对支持MMIO、通道I/O或PCI总线传输外的访客操作系统环境不作假设
- 可扩展 - 设备包含由访客操作系统确定的功能位

Virtio设备设施

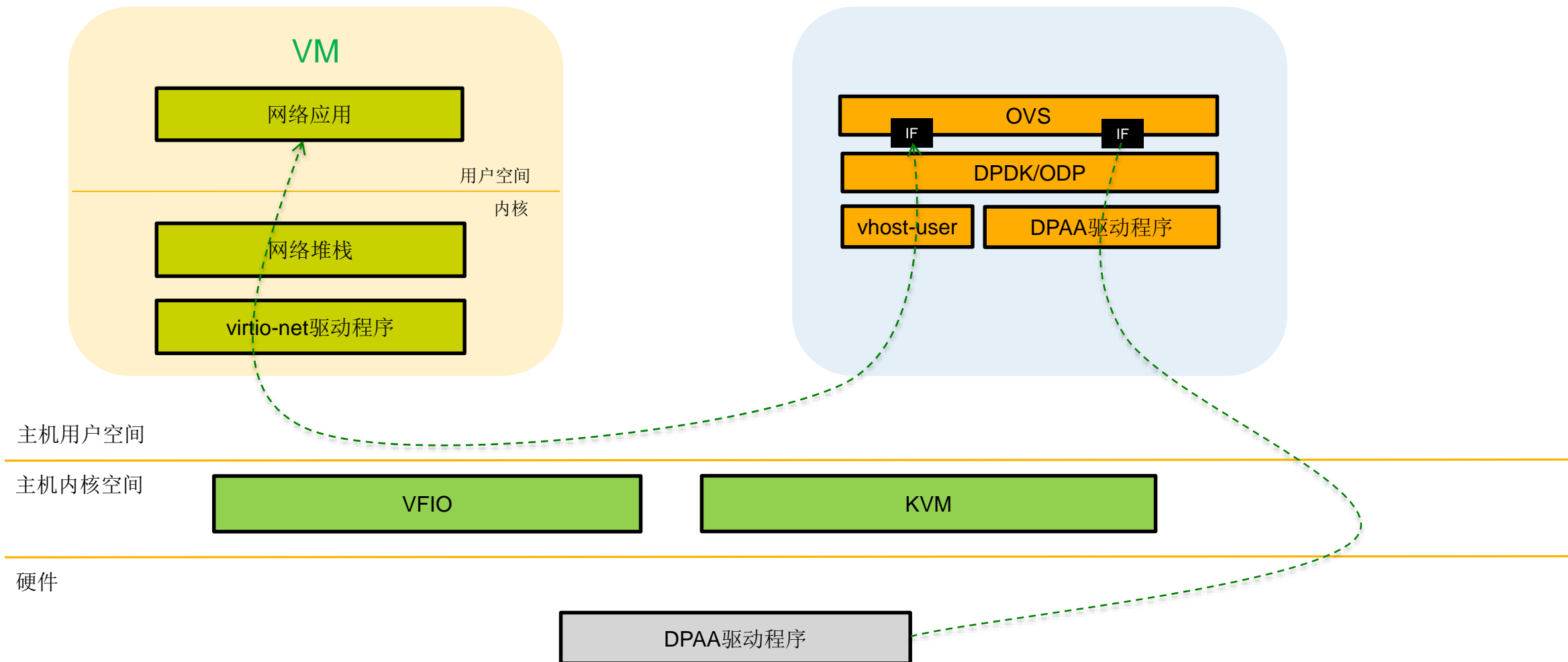
- 设备状态字段
- 功能位
- 设备配置空间
- 一个或多个virtqueue



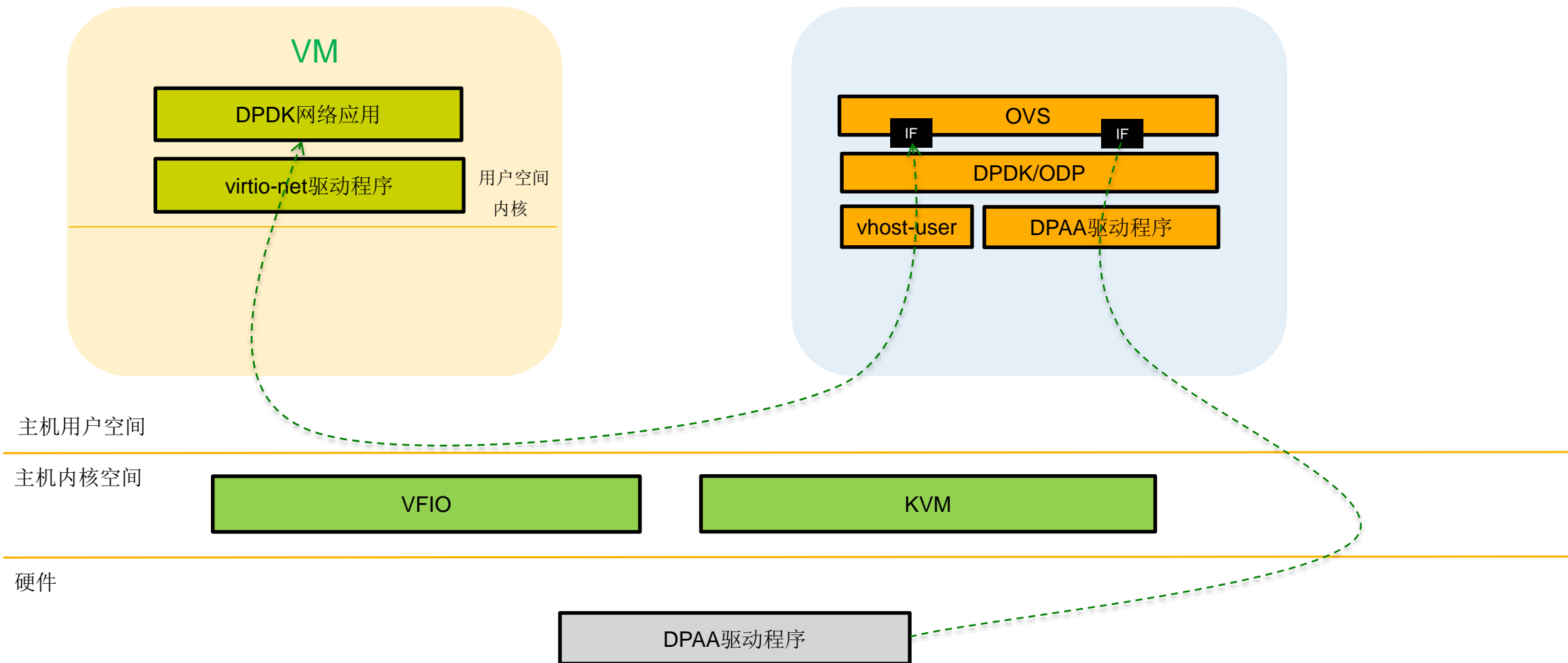
Virtio-net: 用户空间中的Vhost后端



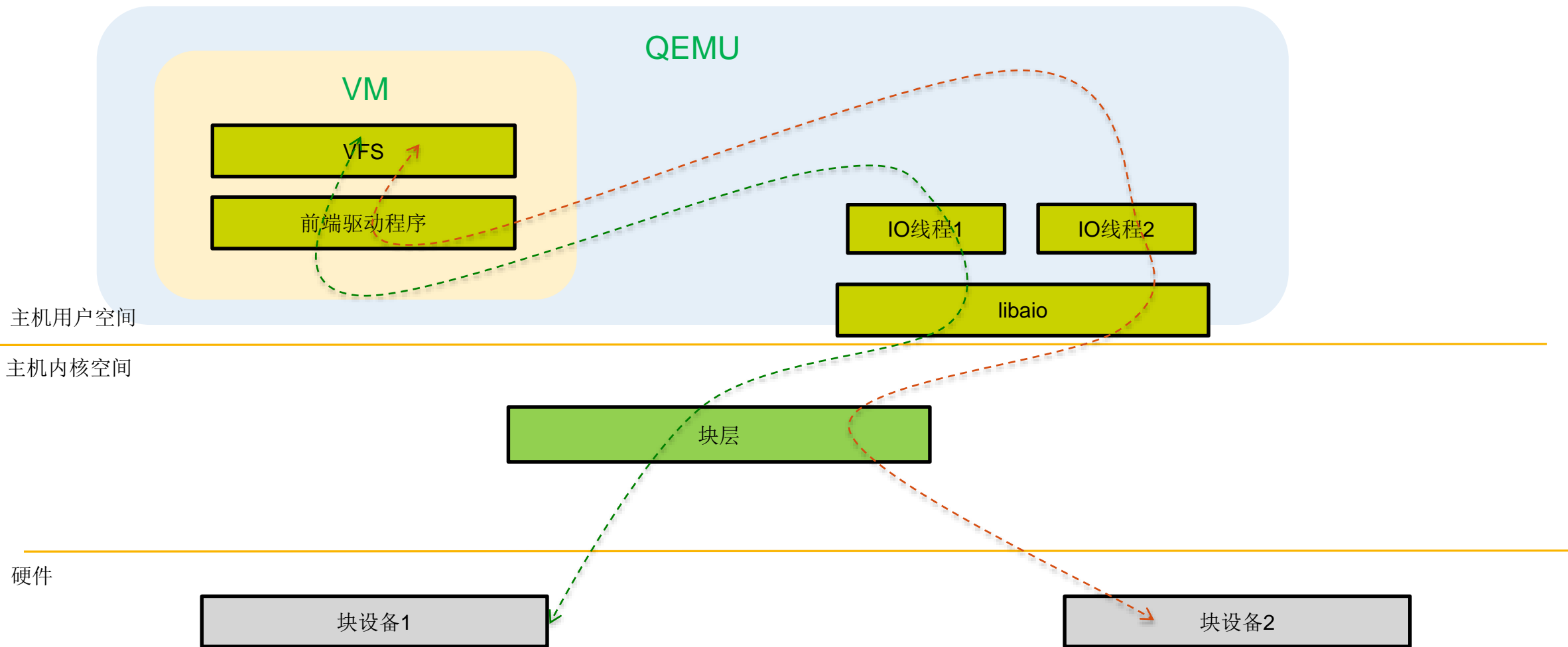
Virtio-net: DPDK-OVS后端



Virtio-net: 使用virtio-net的访客中的DPDK



Virtio-block数据面



结论

结论

- I/O虚拟化解决方案中的效率、性能和灵活性是网络功能虚拟化的重要元素
- KVM提供VirtIO和直接分配，能够让NFV系统设计人员选择最适合其应用的解决方案。



SECURE CONNECTIONS
FOR A SMARTER WORLD

版权声明

恩智浦、恩智浦徽标、恩智浦“智慧生活，安全连结”、CoolFlux、EMBRACE、GREENCHIP、HITAG、I2C BUS、ICODE、JCOP、LIFE VIBES、MIFARE、MIFARE Classic、MIFARE DESFire、MIFARE Plus、MIFARE Flex、MANTIS、MIFARE ULTRALIGHT、MIFARE4MOBILE、MIGLO、NTAG、ROADLINK、SMARTLX、SMARTMX、STARPLUG、TOPFET、TrenchMOS、UCODE、飞思卡尔、飞思卡尔徽标、AltiVec、C 5、CodeTEST、CodeWarrior、ColdFire、ColdFire+、C Ware、高效解决方案徽标、Kinetis、Layerscape、MagniV、mobileGT、PEG、PowerQUICC、Processor Expert、QorIQ、QorIQ Qonverge、Ready Play、SafeAssure、SafeAssure徽标、StarCore、Symphony、VortiQa、Vybrid、Airfast、BeeKit、BeeStack、CoreNet、Flexis、MXC、Platform in a Package、QUICC Engine、SMARTMOS、Tower、TurboLink和UMEMS是NXP B.V.的商标。所有其他产品或服务名称均为其各自所有者的财产。ARM、AMBA、ARM Powered、Artisan、Cortex、Jazelle、Keil、SecurCore、Thumb、TrustZone和 μ Vision是ARM Limited（或其子公司）在欧盟和/或其他地区的注册商标。ARM7、ARM9、ARM11、big.LITTLE、CoreLink、CoreSight、DesignStart、Mali、mbed、NEON、POP、Sensinode、Socrates、ULINK和Versatile是ARM Limited（或其子公司）在欧盟和/或其他地区的商标。保留所有权利。Oracle和Java是Oracle和/或其关联公司的注册商标。Power Architecture和Power.org文字标记、Power和Power.org徽标及相关标记是Power.org的授权商标和服务标记。© 2015–2016 NXP B.V.

