针对DPU的软件无感卸载方案探索

主讲人：李强

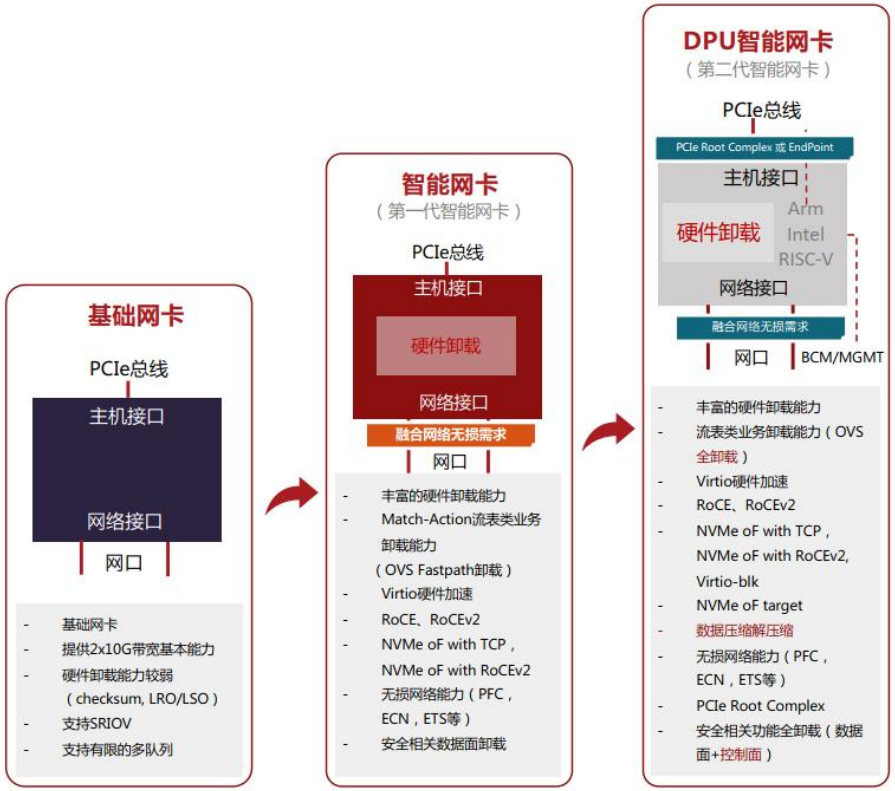
openEuler sig-DPU committer

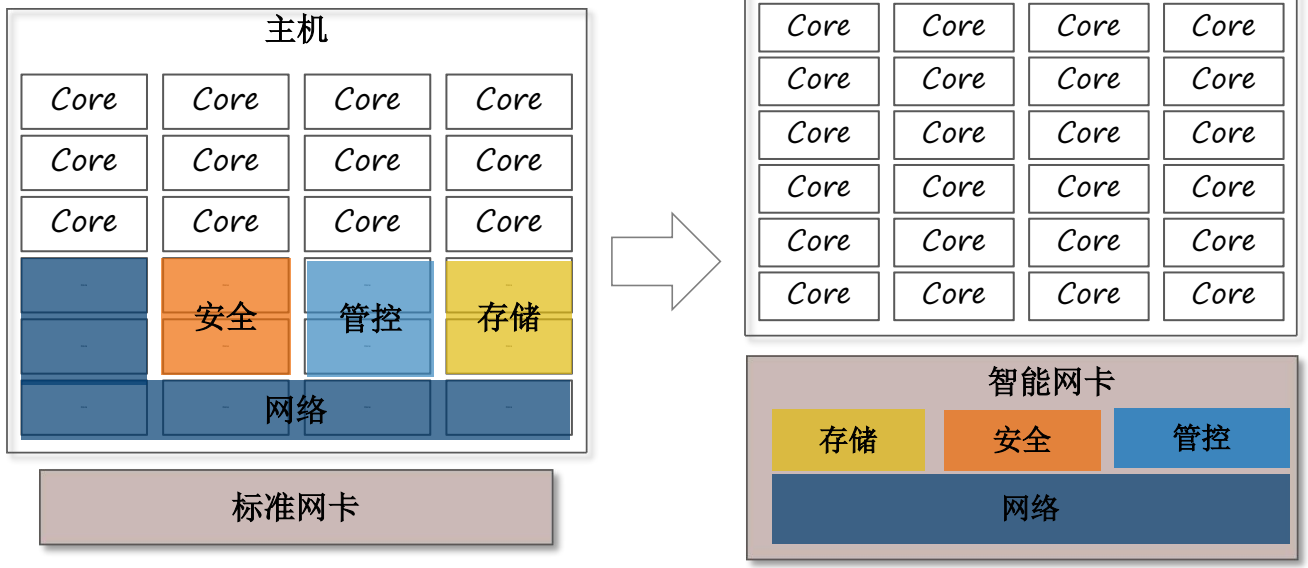
开放原子开源基金会 OPENATOM FOUNDATION | OpenEuler

目录

DPU软件无感卸载的概念和背景

随着传统网卡演进到DPU形态，我们不仅可以用它丰富的处理能力来卸载和加速网络、存储、安全等数据处理相关的厚重协议栈，也可以将一些业务场景的软件管理面组件全卸载到DPU上。





DPU软件无感卸载的概念和背景

管理面软件卸载的难点在于：管理面与业务进程之间深度依赖OS相关机制。

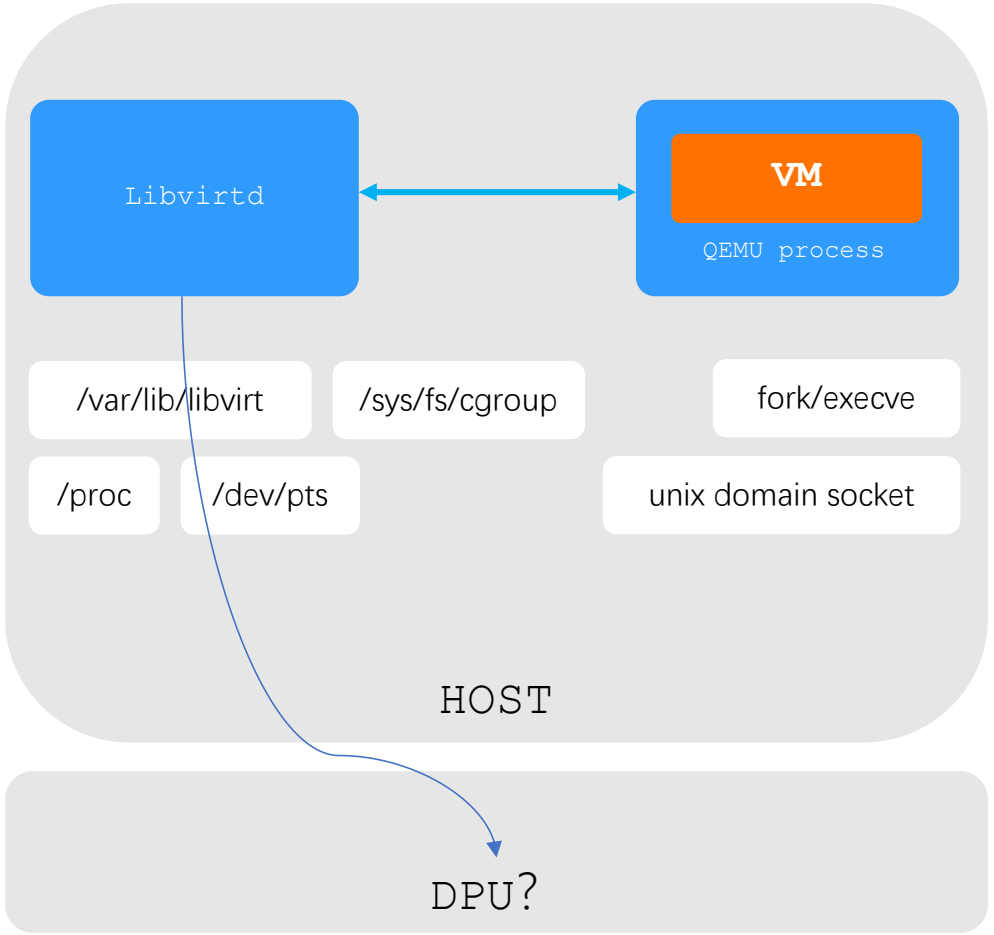
总结起来就是几个大块：

- 文件系统
- 本机IPC（uds、fifo）
- 进程生命周期管理（fork、execve）

如何卸载？

如果将libvirtd组件进行分层拆分，可以将深度依赖OS机制的部分运行在HOST，其他部分运行在DPU。但这需要修改大量源码，并重写Libvirt很多逻辑，且社区无法接纳。

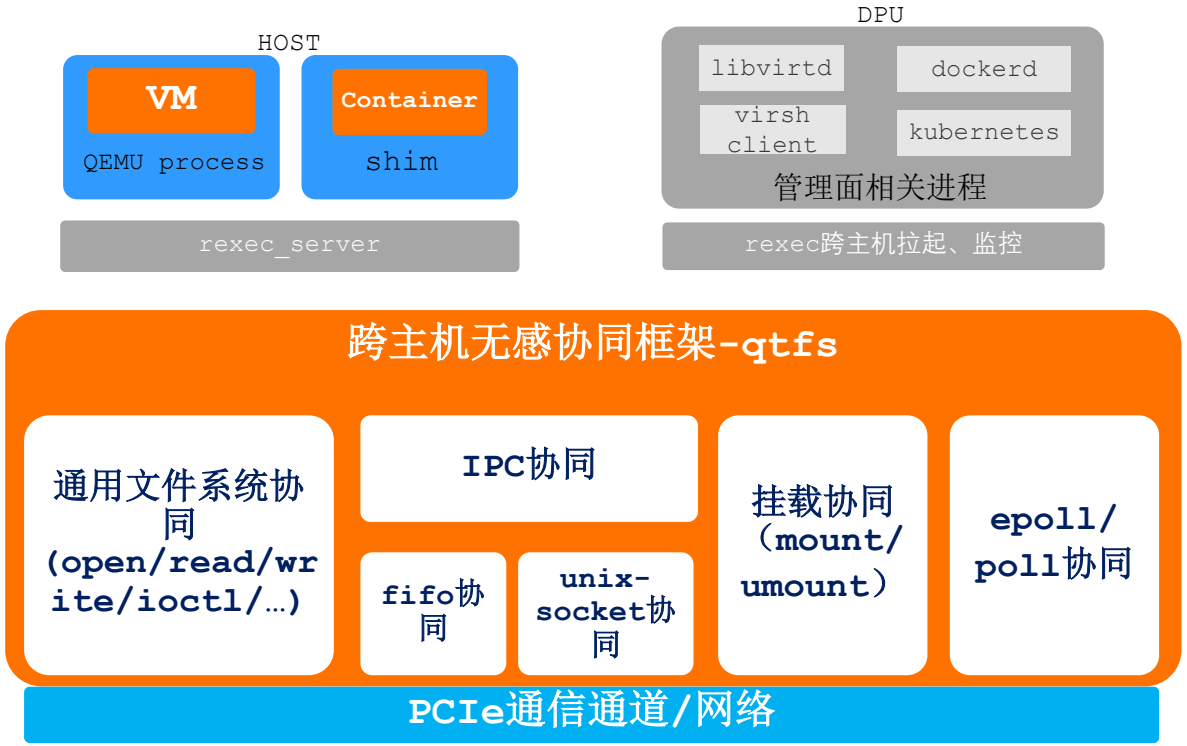
DPU软件无感卸载框架希望探索将兼容层下沉到OS层面，通过聚合HOST-DPU的文件系统、IPC、进程管理等组件，实现libvirt可以全卸载到DPU运行。



DPU软件无感卸载框架实现方案

整体框架

- 通用文件系统聚合：支持HOST-DPU文件系统聚合，为卸载进程提供与业务进程一致的文件系统视图。
- IPC聚合：可以提供完全无感的IPC聚合能力，例如fifo、uds等，为卸载进程与业务进程提供IPC跨机兼容层。
- 进程生命周期管理聚合：管理面进程能使用原生fork+execve的方式管理业务进程在HOST的生命周期。
- 其他聚合能力：挂载协同和epoll协同，卸载进程可以为业务进程部署工作目录并使用epoll等特性。



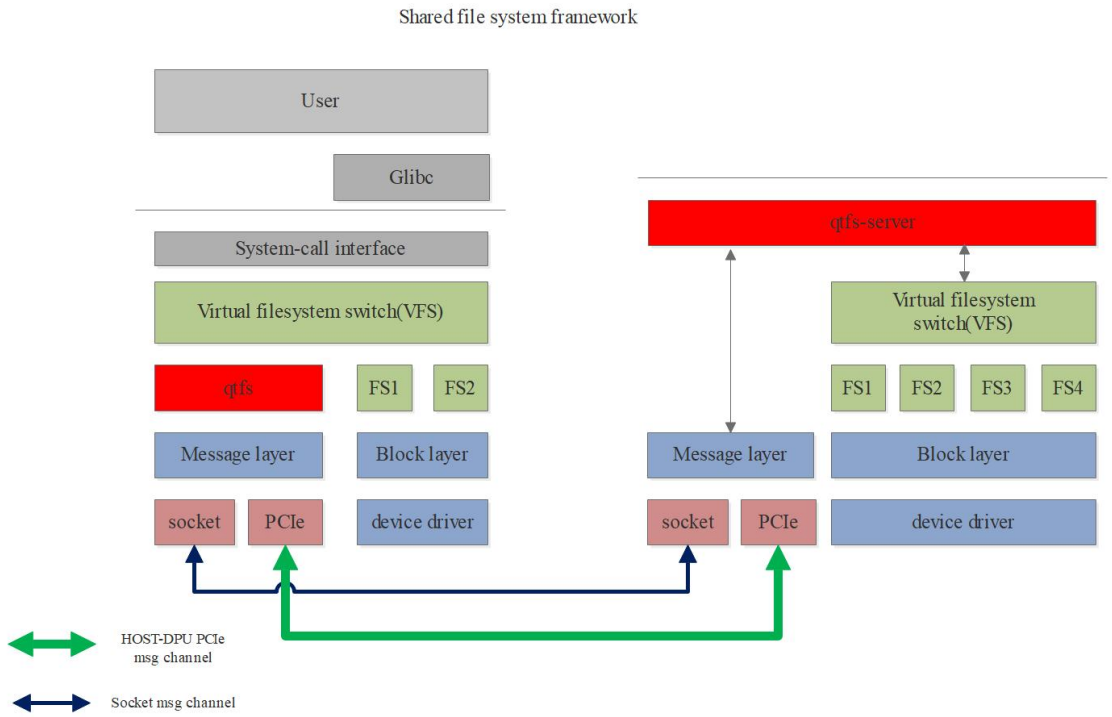
DPU软件无感卸载框架实现方案

OS文件系统协同--qtfs

qtfs是一个共享文件系统项目，可部署在host-dpu的硬件架构上，也可以部署在host-vm或同一台host的vm-vm之间，通过vsock建立安全通信通道。以客户端服务器的模式工作，使客户端能通过qtfs访问服务端的指定文件系统，就像访问本地文件系统一样。

- Client端为qtfs文件系统模块。
- Server端为文件系统代理组件。
- 通过HOST-DPU通信通道（如vsock或其他快速通道）交互

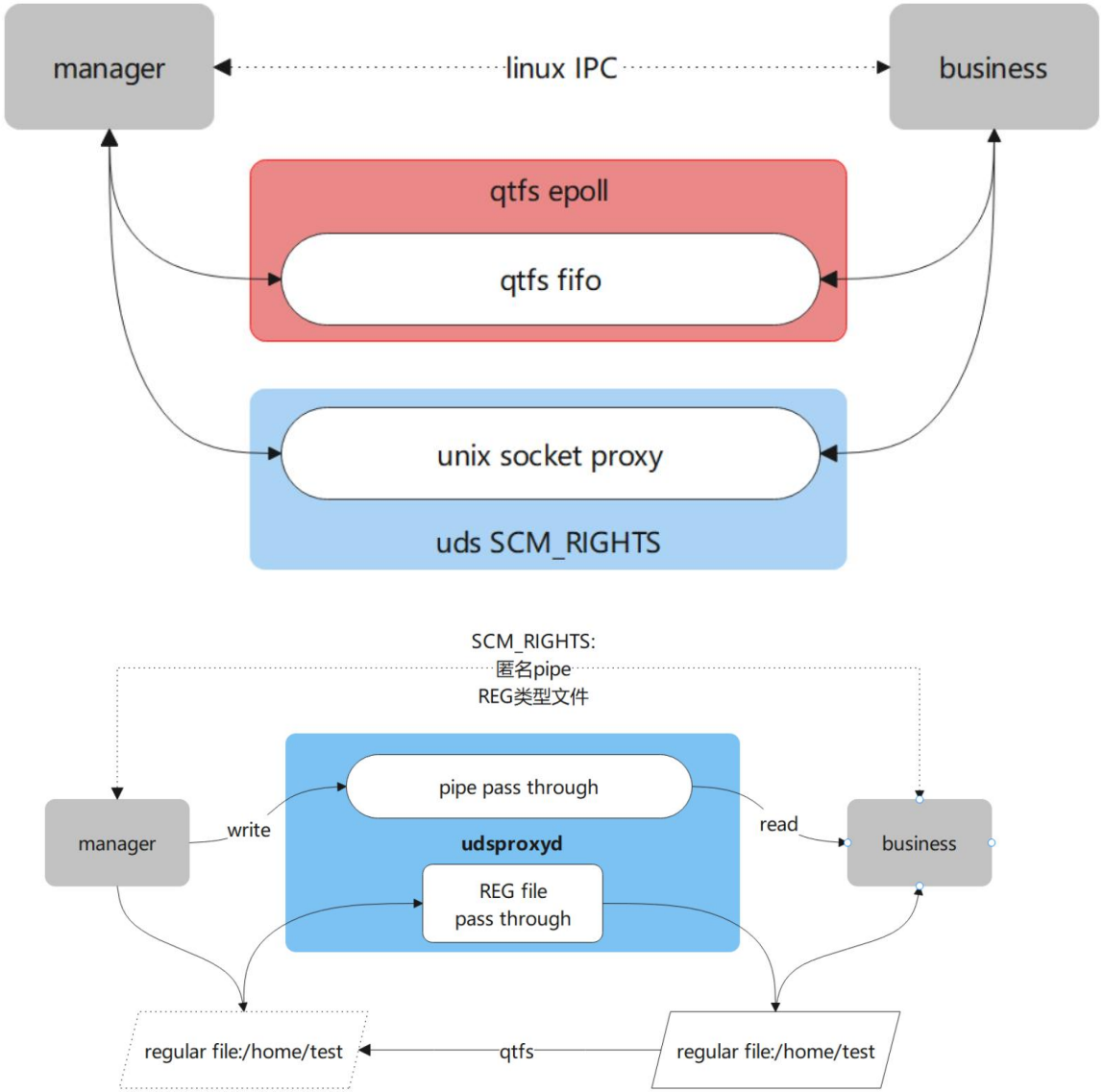
以加速文件系统性能。



DPU软件无感卸载框架实现方案

IPC协同--udsproxy

- Fifo聚合：由文件系统聚合，天然提供了通过fifo文件通信的协同，在容器业务场景，containerd与容器进程之间大量用到fifo通信。
- Uds聚合：由udsproxd守护进程，配合下沉的socket兼容层，实现跨HOST-DPU的unix domain socket聚合。Libvirt主要使用uds与虚拟机进行通信。
- Udsproxyd还能支持跨HOST-DPU进程间的uds SCM_RIGHTS，即父进程将一对pipe的读端或写端传送给子进程，直接读写通信，父子进程都使用标准的linux接口，无需修改。也可以传送REG类型文件给子进程（需要该文件是通过qtfs共享的）。右下图。

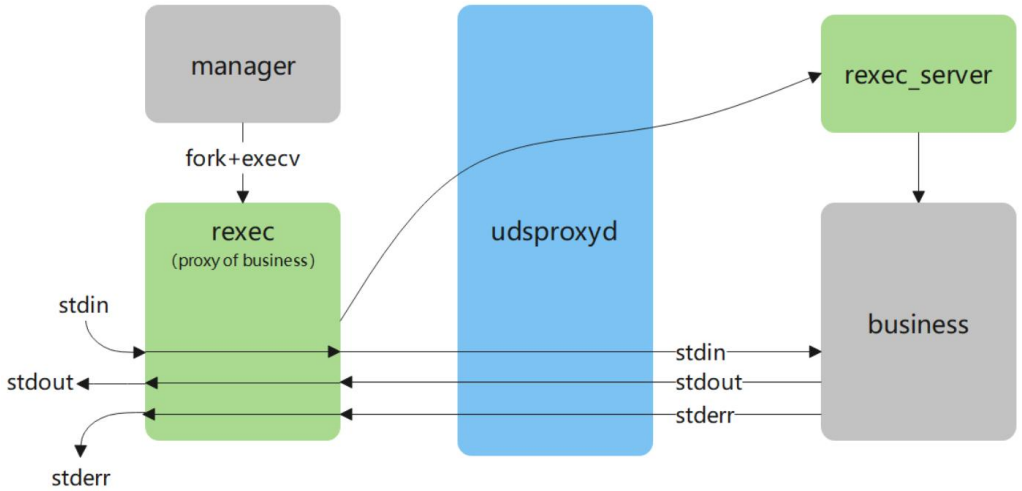


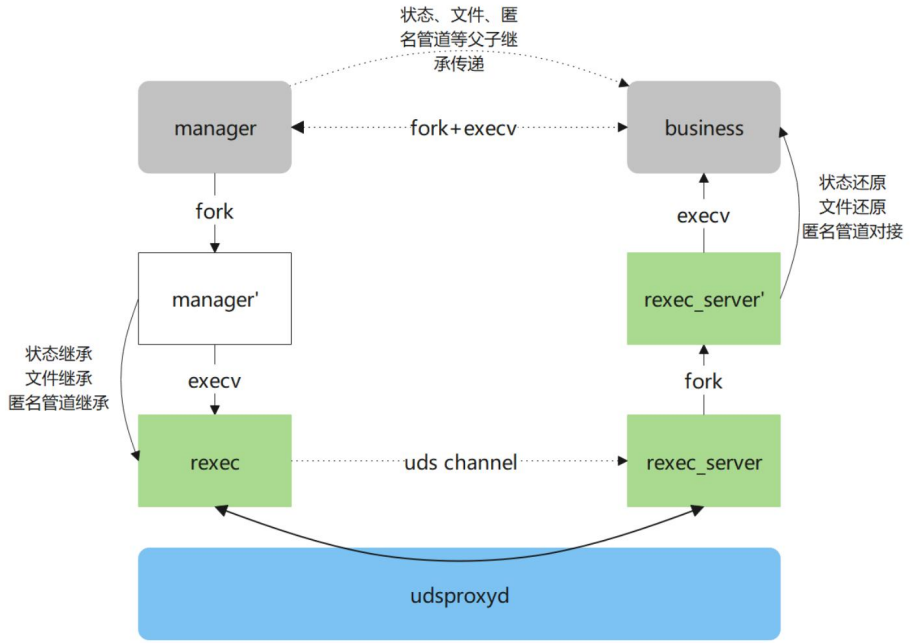
DPU软件无感卸载框架实现方案

业务进程生命周期管理--rexec

Rexec组件在DPU端为一个激活命令rexec，在HOST端为常驻进程Rexec_server，基于udsproxyd组件构建跨主机生命周期管理能力。

- 进程拉起：如右上图，rexec接管新进程拉起，作为远端进程在本地的代理，管理组件通过管理该代理即达到管理远端业务进程的目的。
- 进程状态继承：如下图，rexec把从父进程继承过来的相关状态、匿名管道、打开文件等透传到对端的业务进程，管理进程与在本地拉起业务进程没有太大区别。
- 进程运行时：标准输入输出对接到rexec代理，可在DPU管理侧直接观测到该进程的日志/报错等信息。
- 进程退出：管理进程杀死本地rexec代理时，rexec服务联动杀死对端相应业务进程。

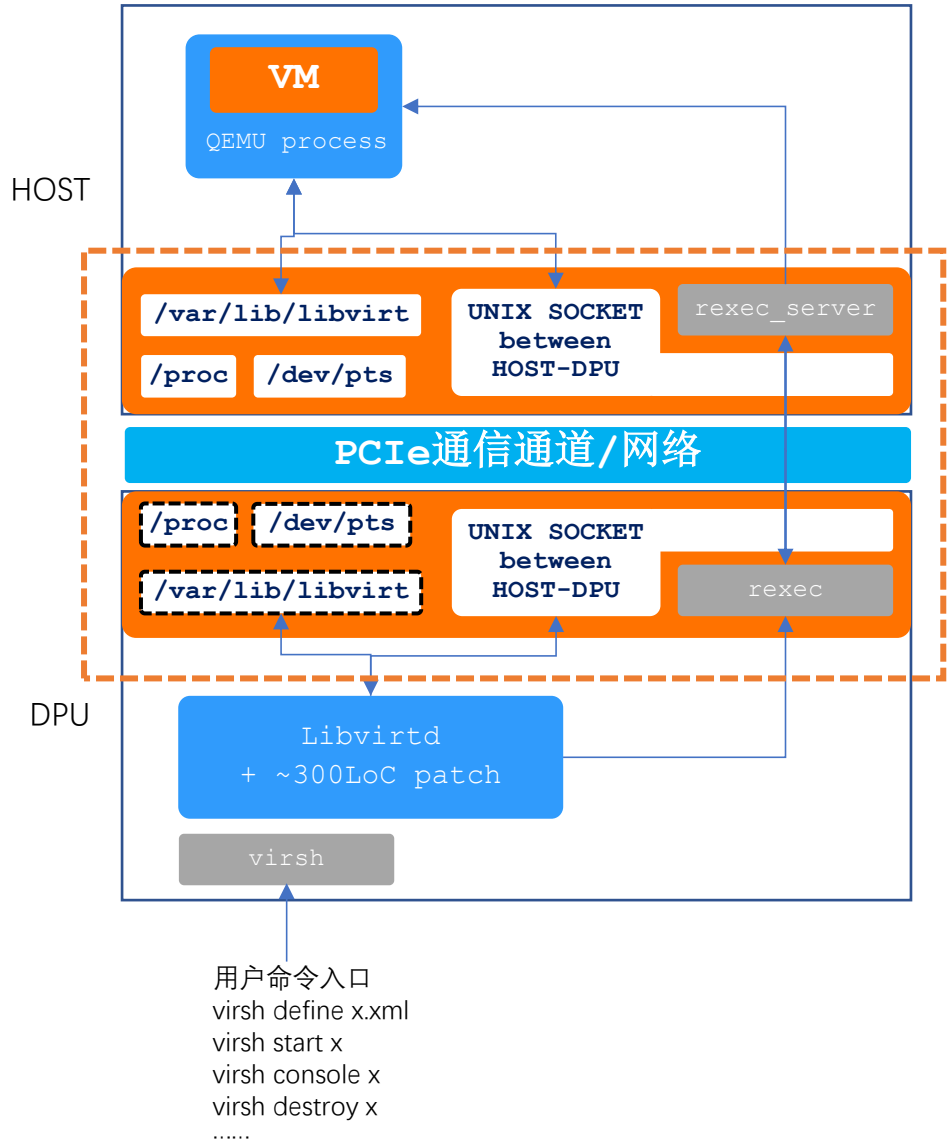




DPU软件无感卸载的应用场景

虚拟化管理面libvirt卸载

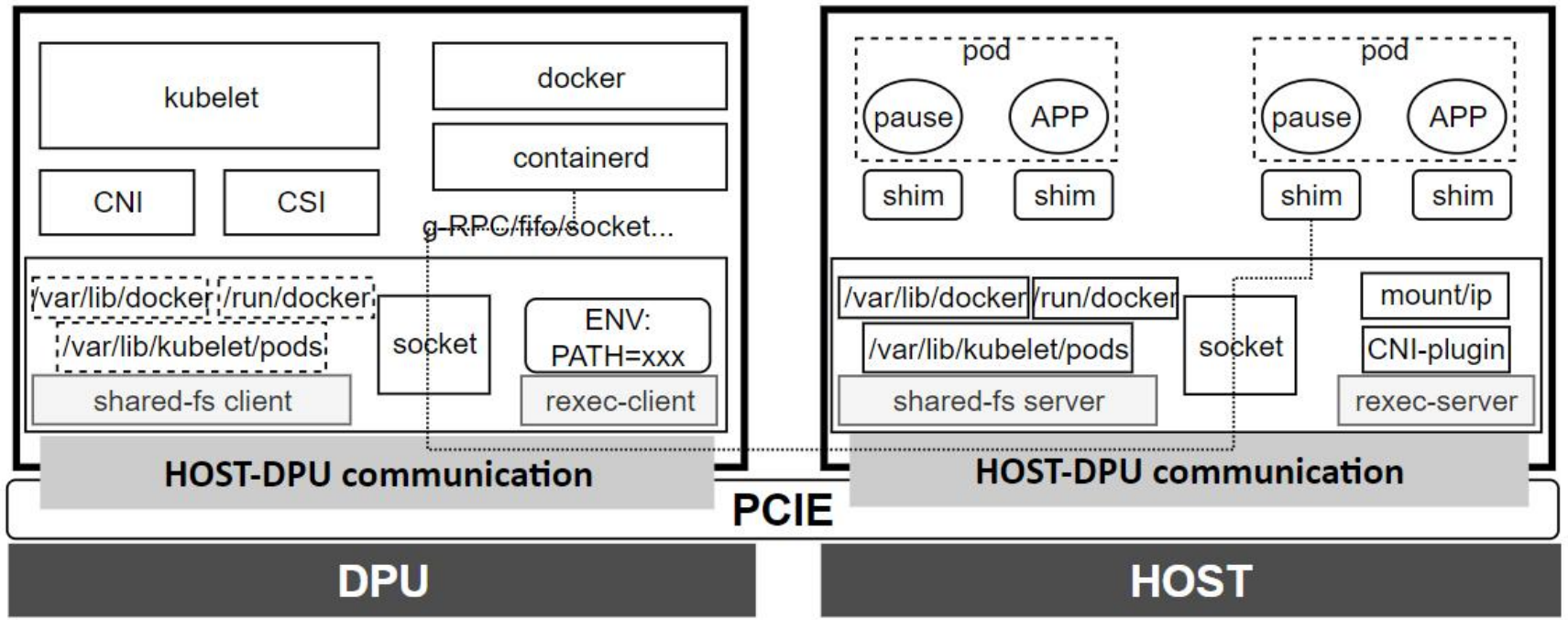
- 目前libvirtd需要做大约300行的适配，在gitee有针对6.2.0版本的patch。
- 无感卸载框架在OS层面，为libvirtd和qemu聚合了前述介绍的文件系统、IPC、进程拉起与管理等，使libvirtd进程能全卸载到DPU。
- 在DPU上virsh命令管理HOST的虚机。
- 全卸载后，HOST的计算资源可以全释放用于运行虚拟机。



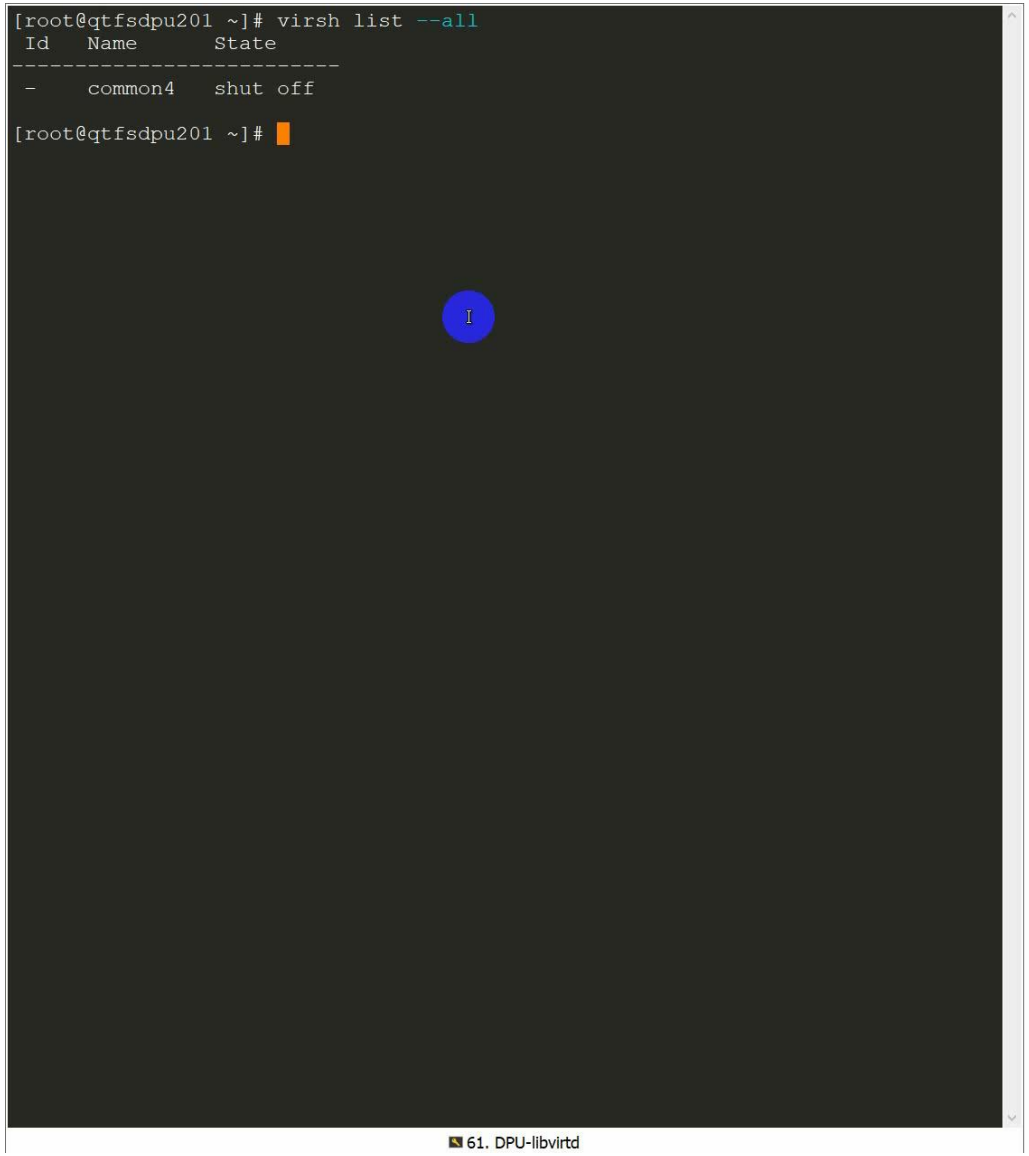
DPU软件无感卸载的应用场景

容器管理面卸载

- 容器管理面也可以使用本组件进行全卸载，其依赖的资源与libvirt类似，无感卸载框架已在OS层面做了通用兼容。



一个演示小demo



项目开源介绍

openEuler源码仓库：https://gitee.com/openeuler/dpu-utilities

src-openEuler包仓库：https://gitee.com/src-openeuler/dpu-utilities

Yum源安装：

openEuler 2203 LTS SP3/SP4、openEuler 2403 LTS：yum install -y qtfs-client qtfs-server

Libvirt卸载相关：https://gitee.com/openeuler/dpu-utilities/blob/master/usecases/libvirtd-offload

环境搭建指南： libvirt直连聚合环境从零搭建v1.1.md

6.2.0适配patch：https://gitee.com/openeuler/dpu-utilities/blob/master/usecases/libvirtd-offload/libvirt-6.2.0/libvirt-6.2.0-offload.patch