Kernel Bypass Technologies Leveraging SmartNICs in the cloud

Presenters: Shachar Raindel, Microsoft Azure

1 INTRODUCTION

This demonstration walks through the architectural changes and the contributions made by Microsoft to several open source technologies, so that kernel bypass technologies can be "cloud ready". We leverage the SmartNICs that are deployed in Azure to provide this functionality [1].

Recently, there is a huge momentum behind kernel bypass technologies to enable workloads targeting ultra-low latencies and millions of packets per second (PPS). These kernel bypass technologies take a direct dependency on the underlying hardware like the network adapter, challenging existing virtualization solutions. As these workloads move to the cloud, it is very important to make sure these kernel bypass technologies can move to the cloud and continue to function with high performance and reliability. Specifically, we will demonstrate two key properties enabled by our SmartNIC innovation.

a) Serviceability in the cloud

One of the key aspects of running the cloud is the high pace of innovation and scale improvements in the underlying platform. This make serviceability one of the most important pillars of the cloud.

This requirement was relatively straightforward for traditional networking model -- it is constrained to use synthetic devices in the VM, through a cloud vendor provided device driver in the VM OS. There is no direct interaction between the VM and the physical server HW being serviced. However, such model suffers from poor performance.

To achieve higher performance, recent workloads turn to SR-IOV accelerated networking. However, this exposes the underlying device directly into the VM. Kernel bypass in the VM stack implies the application is now directly interacting with the queues of the underlying physical hardware. When the host components are serviced, the corresponding device is revoked from the VM, breaking the kernel bypassing path to the network.

To solve this problem, *i.e.,* to ensure the applications stay up and running while host components are serviced, we implement a fallback to a synthetic network device. This fallback is performed transparently to the application, while kernel bypass is enabled and the application is interacting directly with the underlying queues. This work went into DPDK (Data Plane Development Kit) [2]. DPDK is one of the major kernel bypass solutions. It is being used by Network Appliances to process millions of PPS. All of the work done for Azure has been contributed back to DPDK upstream and merged to the 18.05 version, enabling other platforms can make use of the engineering effort and architecture.

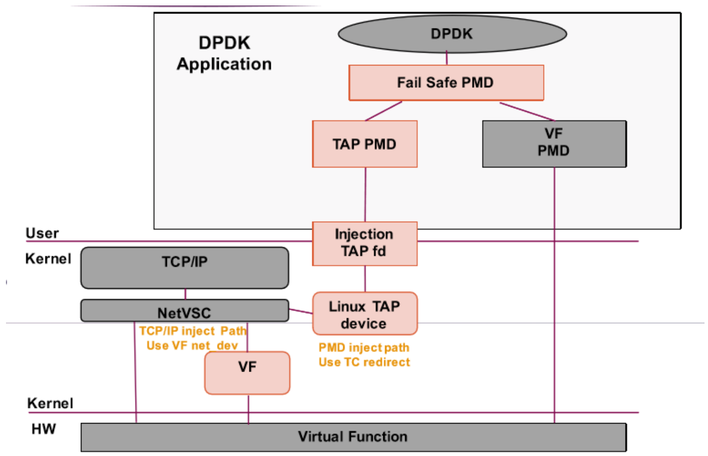

Figure 1: DPDK Architecture with Fallback Support

b) Lowest latnecy in the cloud

In addition to DPDK, we want to enable also legacy socket applications to leverage kernel bypassing acceleration to reduce messaging latency. We enhanced VMA (Messaging Accelerator) [3] to be compatible with cloud environment. As part of this demonstration, we will also show the lowest latencies achievable in any cloud environment today using pure socket applications. All of these scenarios make use of the SmartNICs to offload all SDN processing to FPGAs.

2 Demo Details

We will show virtual machines running on Azure pre-prod environment with the technologies abovementioned enabled. The VMs will perform BW and latency benchmarks using the frameworks listed. We will demonstrate the serviceability features by forcing removal of the VF and bringing it back, with customer load running uninterrupted.

3 Summary

Microsoft Azure is making huge investments in ultra low latency and high PPS scenarios and contributing this technology back to the open source world so that it can be used to onboard excited new scenarios from researchers and various other industries. Researchers can leverage Azure to prototype and benchmark innovative networking systems. The accelerated networking and kernel-bypassing support in Azure enables the research community to achieve realistic results for system performance with minimal HW investment. Azure is also committed to open-source and academic collaboration.

References:

[1] Azure Accelerated Networking: SmartNICs in the Public Cloud, Daniel Firestone et al., 2018 https://www.usenix.org/system/files/conference/nsdi18/nsdi18-firestone.pdf

[2] DPDK: Data Plane Development Kit. http://dpdk.org/about , 2018.

[3] VMA: Messaging Accelerator. http://www.mellanox.com/page/software_vma?mtag=vma , 2018