NETRONOME

WHITE PAPER

The Case for Express Virtio (XVIO)

XVIO INCREASES THE OPERATIONAL EFFICIENCY OF OPENSTACK CLOUD SERVER INFRASTRUCTURES

DELIVERING OPTIMAL I/O BANDWIDTH TO VMS OF DIFFERENT FLAVORS IS A CHALLENGE; IT REQUIRES DIFFERENT NETWORKING CONFIGURATIONS AND AFFECTS THE HETEROGENEITY AND EFFICIENCY OF SERVER INFRASTRUCTURES..

INTRODUCTION

Virtual machines (VMs) and the applications that run in them form the core building blocks of today's modern cloud data centers, and revenue from VMs is the lifeline of public cloud service providers. Enterprises that deploy private cloud infrastructure rely on VMs and their apps to deliver the needed services to their customers. Data center operators rely on the promise of deploying SDN and NFV, and realizing deployment efficiencies using COTS servers to achieve their revenue, service levels and other business goals. However, all of them echo what James Hamilton of Amazon Web Services said a few years ago, "Data center networks are in my way," but with a subtle difference. Instead of data center networks, some said SDN, and others said OpenStack networking or Open vSwitch (OVS) networking or Contrail vRouter networking were what was in there way. When it came to describing their challenges, the resounding common theme was poor server infrastructure efficiency. It sure hurts when 60% of data center infrastructure costs come from servers and the cost to power and cool them.

So what is it about current networking solutions that is really hurting OpenStack-based server infrastructure efficiencies? Let's look under the hood.

VM WORKLOADS DIFFER – EACH REQUIRES A DIFFERENT NETWORKING CONFIGURATION

VMs come in many flavors based on what they run: web server applications, Hadoop big data or other database applications, networking applications, or security applications. The VMs have different profiles in terms of resource requirements: the number of virtual CPU cores (vCPUs), the amount of I/O bandwidth (packets per second), memory (megabytes), disk (gigabytes), tenancy requirements (single versus multiple), reporting of metrics related to their operation and performance behavior, and others. The vCPUs and amount of I/O bandwidth (packets per second) for each VM are relevant in this discussion. For example, a web server application typically requires a moderate number of vCPU cores and lower I/O bandwidth, while Hadoop big data and networking application VMs typically require a moderate-to-high number of vCPU cores and high I/O bandwidth. Delivering optimal I/O bandwidth to VMs of different flavors is a challenge; it requires different networking configurations and affects the efficiency of server infrastructures.

NETWORKING CONFIGURATION OPTIONS: VIRTIO, DPDK OR SR-IOV

Networking services for VMs in OpenStack deployments are mostly delivered using OVS or Contrail vRouter. The I/O bandwidth delivered through such datapaths to VMs is affected by the following:

- Datapath decisions made on the host or hypervisor. The OVS and vRouter datapaths that deliver needed network services can execute in one of three modes on the host or hypervisor:
  1. **Linux Kernel Space:** The datapath typically executes in the Linux kernel space running on x86 CPU cores. This mode delivers the lowest I/O bandwidth, consuming a high number of CPU cores. More CPU cores can be allocated to the kernel datapath processing to enable higher performance, but in most cases performance caps off at less than 5 Mpps (million packets per second) with about 12 CPU cores. The network administrator has to pin CPU cores to the datapath processing tasks to ensure predictable performance while blocks these cores from being utilized by VMs and applications..
  2. **Software Acceleration (DPDK):** For higher performance, the datapath may be executed in the Linux user space using the Data Plane Development Kit (DPDK) running on x86 CPU cores. More CPU cores can be allocated to the user space datapath processing to enable higher performance, but in most cases performance caps off at less than 8 Mpps with about 12 CPU cores. Again, the network administrator has to pin CPU cores to the datapath processing tasks to ensure predictable performance.
  3. **Hardware Assisted Bypass (PCIe Passthrough or SR-IOV):** In case of PCIe Passthrough the entire PCIe device is mapped to the VM bypassing the hypervisor. In case of SR-IOV the hardware and the firmware provide a mechanism to segment the hardware and be used by multiple VMs at the same time. SR-IOV is the technology that is relevant to virtualization in this context.
  4. **Hardware Acceleration (SmartNIC):** For even higher performance, the datapaths may execute in a SmartNIC such as Netronome's Agilio platform, which can achieve up to 28 Mpps of performance, consuming only one CPU core for control plane processing related to the OVS or vRouter datapath. Since the datapath runs in dedicated network processing cores in the SmartNIC, performance is predictable and no extra provisioning tasks are required from the network administrator.
- The I/O Interface between the VM and Host or Hypervisor. The data can be delivered to the VM from the OVS and vRouter datapaths in one of the following ways:
  1. **Virtio:** In this case, VMs are completely hardware independent and can therefore be easily migrated across servers to boost server infrastructure efficiency. Applications running in all popular guest operating systems in the VMs require no change, making onboarding of customer or third party VMs easy. Live migration of VMs across servers is feasible. Networking services provided by the OVS and vRouter datapaths are available to the VMs, however I/O bandwidth to and from the VM is lower than with DPDK and SR-IOV.
  2. **DPDK:** In this case, VMs require a DPDK poll mode driver that is hardware independent. Applications need to be modified to leverage the performance benefits of DPDK, and as a result, customer and third party VM and applications onboarding is not as seam-

less. Live migration of VMs across servers is feasible. Networking services provided by the OVS and vRouter datapaths are available to the VMs, but in a limited way if the user space DPDK datapath is used because such services typically evolve rapidly in the kernel and need to be ported and made available in the user space, and this may take time or may be difficult to implement. (This deficiency of DPDK was underscored by the Linux kernel maintainer David Miller at the recent Netdev1.2 conference in Tokyo, when he very aptly repeated multiple times that "DPDK is not Linux.") I/O bandwidth to and from the VM is higher than Virtio but significantly lower than Single Root I/O Virtualization (SR-IOV).

3. **SR-IOV (Single Root I/O Virtualization):** In this case, VMs require a hardware-dependent driver in the VM, but applications do not need to be modified to leverage the performance benefits of SR-IOV. Customer and third party VM and applications onboarding is impossible unless the vendor hardware driver is available in the guest operating system. Live migration of VMs across servers is not feasible because drivers are hardware dependent. Networking services provided by the OVS and vRouter datapaths are not available to the VMs if they are implemented in the kernel space or user space with DPDK. Networking services provided by the OVS and vRouter datapaths are not available to the VMs if they are implemented in a SmartNIC. I/O bandwidth to and from the VM is the highest using SR-IOV.

WHEN OPERATORS HAVE TO DEAL WITH DIFFERENT VM PROFILES, IT IS IMPOSSIBLE TO ACHIEVE AN OPTIMAL, HOMOGENOUS END-TO-END CONFIGURATION ACROSS ALL SERVERS.

ADVERSE EFFECTS ON DATA CENTER OPERATIONAL EFFICIENCY

When operators have to deal with different VM profiles, it is impossible to achieve an optimal, homogenous end-to-end configuration across all servers. For example, the operator could take different paths based on data path options in the host/hypervisor and the I/O interface between the VM and the host/hypervisor:

OVS in Host + Virtio in VM: One option could be to use the kernel datapath options for OVS or vRouter and Virtio-based delivery of data to VMs across all servers. The result is a homogenous server deployment managed using OpenStack. The challenge here is poor performance to the VMs that require higher packets per second, or not having enough CPU cores left to deploy an adequate number of VMs in the server. This results in poor SLAs and server sprawl.

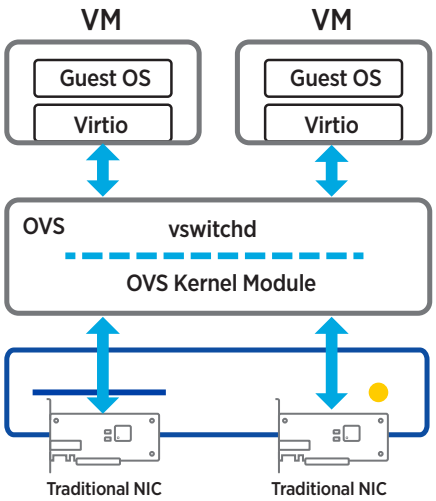

Figure 1. OVS in Host + Virtio in VM.

OVS-DPDK in Host + DPDK PMD or Virtio to VM: A second option is to use DPDK user space datapath options for OVS or vRouter and DPDK and Virtio-based delivery of data to VMs across all servers. The result is a homogenous server deployment managed with OpenStack. The challenge here is mediocre performance to the VMs that require higher packets per second, or not having enough CPU cores left to deploy an adequate number of VMs in the server. Also, to deliver adequate levels of performance, the administrator will have to pin different numbers of cores to the DPDK datapath to get the right level of performance. As a result, cores could be wasted on some servers or performance may be inadequate on others. Enabling an efficient way of configuring DPDK and the right number of cores per server can become a nightmare, taking operational costs higher. This challenge is further explained in the next section.

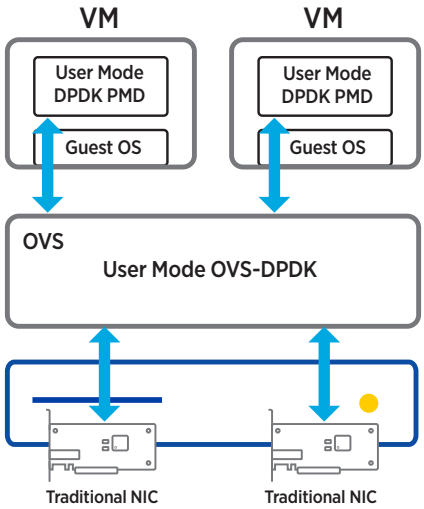

Figure 2. OVS-DPDK in Host + DPDK PMD or Virtio to VM.

Traditional NIC + SR-IOV: A third option is to create a silo of servers that are configured using SR-IOV to deliver the highest performance. If the datapaths are running in kernel or user space, as discussed earlier, all SDN-based services provided by OVS or vRouter are lost. VMs on this silo of servers cannot be migrated. This makes efficient management of the servers difficult, taking operational costs higher.

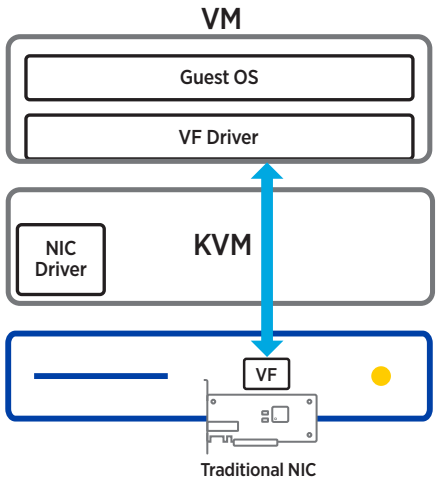

Figure 3. Traditional NIC + SR-IOV.

SmartNIC + SR-IOV: A fourth option is to create a silo of servers that are configured using SR-IOV to deliver the highest performance. The datapath runs in a SmartNIC, as discussed earlier, and all SDN-based services provided by OVS or vRouter are kept intact. VMs on this silo of servers cannot be migrated, however. This makes efficient management of the servers difficult, taking operational costs higher.

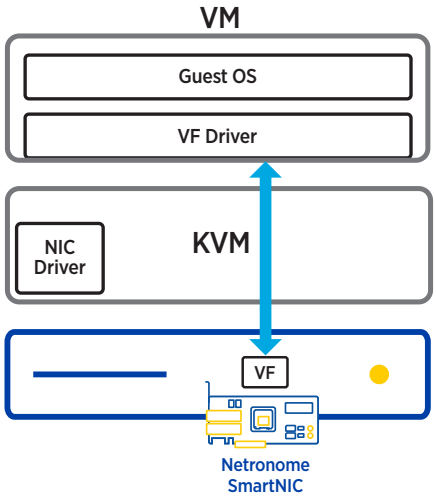

Figure 4. SmartNIC + SR-IOV.

As can be seen, none of the above options meets the performance, networking options and resource utilization requirements of the VMs. The data center operators end up selecting multiple options, which ultimately means a heterogeneous environment that has high CAPEX and OPEX, and reduced utilization. This is depicted in Figure 5. As can be seen, none of the above options holistically meets the performance, flexibility, networking services and resource utilization requirements of the VMs and the servers they are hosted in.

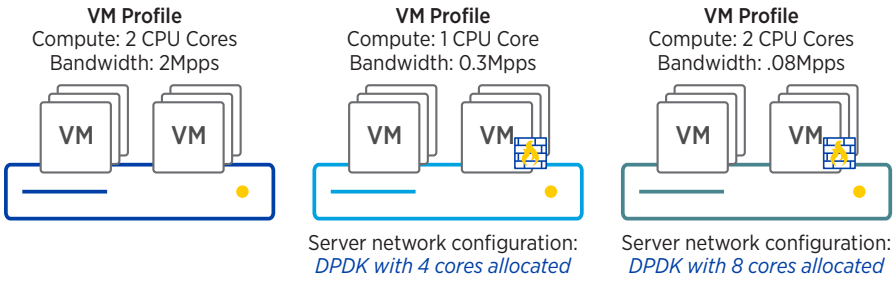

Figure 5. Servers are configured differently to service different VM profiles. VMs cannot be migrated at all or it is hard to do so.

AN INDUSTRY FIRST, NETRONOME'S INNOVATIVE EXPRESS VIRTIO (XVIO) TECHNOLOGY ELIMINATES SIGNIFICANT OPERATIONAL, PERFORMANCE AND SERVER EFFICIENCY-RELATED CHALLENGES.

DPDK: ROBBING PETER TO PAY PAUL

OVS or vRouter datapaths implemented in the user space using DPDK have provided some relief for VM profiles that require more I/O bandwidth. DPDK proponents tout DPDK as a method for solving performance bottlenecks in the world of NFV. However, since applications running in VMs (virtual network functions or VNFs, for example) have to share the same available CPU cores on the server with the CPU cores that need to be allocated to run the OVS or vRouter datapaths implemented in the user space using DPDK, one quickly runs into the "robbing Peter to pay Paul" scenario. For example, to service VM profiles of one kind, one may allocate eight cores to DPDK OVS or vRouter; for another VM profile, one may allocate four cores to DPDK OVS or vRouter; and for another VM profile, one may allocate 12 cores to DPDK OVS or vRouter. In some scenarios, the VM profile that needs DPDK OVS or vRouter with 12 cores most likely also needs the largest number of CPU cores for application performance. Efficient distribution of cores becomes a challenge and this is further exacerbated when one has a mix of VM profiles on the same server, some requiring higher CPU cores than others and some requiring lower bandwidth that others.

EXPRESS VIRTIO (XVIO) TO THE RESCUE

An industry first, Netronome's innovative Express Virtio (XVIO) technology eliminates the significant operational, performance and server efficiency-related challenges highlighted above. XVIO brings the level of performance of SR-IOV solutions to standard Virtio drivers (available in many guest OSs) but maintains full flexibility in terms of VM mobility and provides the full gamut of network services provided by OVS and vRouter. This enables VMs managed using OpenStack to experience SR-IOV-like networking performance while at the same time supporting complete the hardware independence and seamless customer VM onboarding of Virtio. For the cloud service provider, the benefit of utilizing OpenStack cloud orchestration is a consistent and homogenous infrastructure where VMs can be placed and moved to optimize utilization of the data center while maintaining high performance.

The following figures illustrate the operational efficiencies that XVIO brings for SDN-based data center infrastructure deployments.

Figure 6 shows how XVIO fares versus VM data delivery mechanisms such as DPDK and SR-IOV when the OVS or vRouter datapaths are implemented in the Netronome Agilio SmartNICs and server networking platform. In this figure, flexibility in terms of customer VM onboarding

and live VM migration is mapped versus performance delivered to VMs.

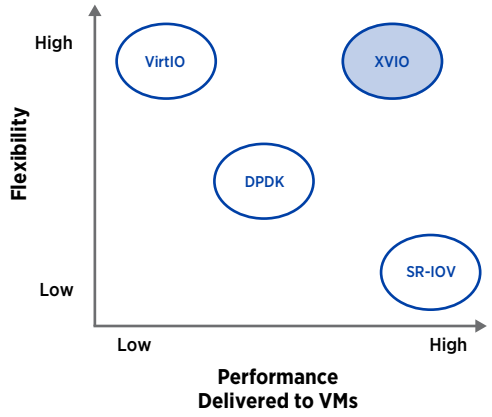

Figure 6. Highly Flexible XVIO - rapid customer VM onboarding, live VM migration.

Figure 7 shows how XVIO fares versus VM data delivery mechanisms such as DPDK and SR-IOV when the OVS or vRouter datapaths are implemented in the Netronome Agilio SmartNICs and server networking platform. In this figure, rich SDN-based features such as policy rules with ACLs or security groups, flow-based analytics, or load balancing are mapped versus performance delivered to VMs.

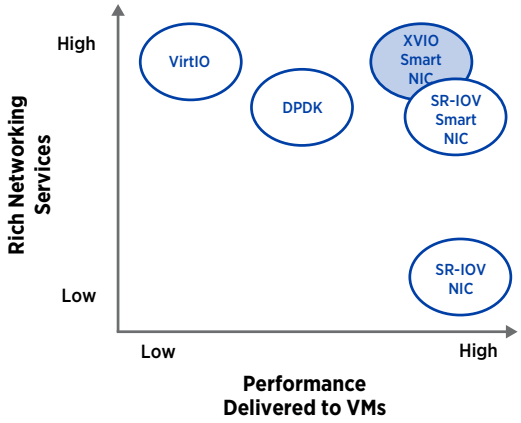

Figure 7. Rich Networking Services with XVIO - policy rules with ACLs, flow-based analytics, and load balancing.

Figure 8 shows how XVIO fares versus VM data delivery mechanisms such as DPDK and SR-IOV when the OVS or vRouter datapaths are implemented in the Netronome Agilio SmartNICs and server networking platform. In this figure, high server efficiency metrics such as freeing up CPU cores for applications and VMs are mapped versus performance delivered to VMs.

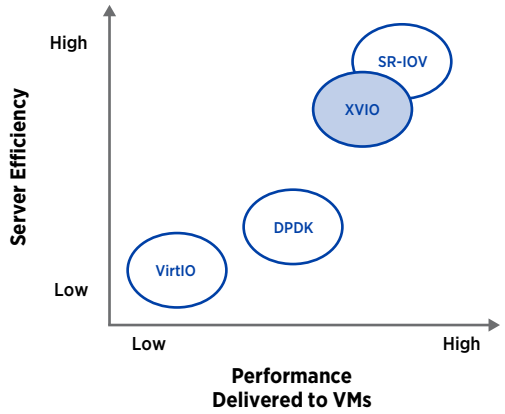

Figure 8. High Server Efficiency with XVIO - freeing up CPU cores for applications and VMs.

The advanced XVIO technology is based on and builds upon industry standard and open source technologies such as SR-IOV, Virtio and DPDK supported by OpenStack. The XVIO technology and software components are transparent, and integrate easily with open source and commercial server networking software such as OVS, Linux Firewall and Contrail vRouter. VMs and their applications do not require any changes, and all popular guest operating systems with standard Virtio drivers are supported.

XVIO implemented in the Netronome Agilio server networking platform reduces operational complexity. For the cloud service provider, the benefit of utilizing OpenStack cloud orchestration is a consistent and homogenous infrastructure where VMs can be placed and moved to optimize utilization of the data center while maintaining high performance. This is depicted in Figure 9.

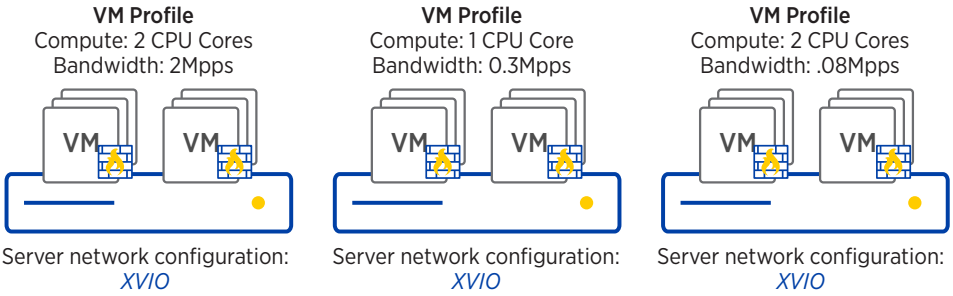

Figure 9. With XVIO, servers are configured the same way; VMs with different profiles can be migrated and placed most efficiently.

SUMMARY

Private and public cloud deployments use SDN and cloud-based orchestration based on OpenStack or operator-developed centralized SDN controllers. They leverage networking and security services delivered by OVS and Contrail vRouter that run in servers. To enable virtualized server-based network performance scaling in cloud deployments, the industry has employed a number of acceleration mechanisms. DPDK requires changes to applications and VMs (adversely affecting server efficiency) and cannot use key Linux kernel-based networking

services. Unlike SR-IOV, a PCI-SIG technology, XVIO does not limit VM mobility and availability of networking and security services needed by VMs, but delivers the same level of performance.

Netronome Agilio server networking platforms with XVIO technology deliver a simple deployment model that removes barriers and makes adoption of networking accelerators such as SmartNICs economical and practical, saving CAPEX and OPEX in a significant way. In short, XVIO with the Netronome Agilio server networking platform makes OpenStack, and in general cloud networking, faster and more economical.

NETRONOME

Netronome Systems, Inc.
2903 Bunker Hill Lane, Suite 150 Santa Clara, CA 95054
Tel: 408.496.0022 | Fax: 408.586.0002
www.netronome.com