Emergence of programmable SmartNICs for the Edge

Rhod Brown

2019









© 2019 Mellanox Technologies

Where we have been, to where we are now





2007 Onwards... Network virtualization approach: Software vendors Virtual applications Standard high-volume servers Standard high-volume storage Standard high-volume Ethernet switches

Software Defined Everything World

© 2019 Mellanox Technologies

NFV And Networking

Networking plays significant Role in NFV transformation





© 2019 Mellanox Technologies

NFV Transformation Challenges

Economics Viability

- Commercial Off The Shelf (COTS) HW should be at similar scale as proprietary one
- For this COTS networking and processing performance should be adequate
- Better, faster, accelerated infrastructure is needed
 - While maintaining standards and available off the shelf
- Programmability, SDX
 - The main concept of cloud is agility and automation
 - Software Defined Everything
 - NFV cloud have stronger requirements on network programmability than any other cloud use case
 - Flexible control and data path is mandatory







Network functions virtualization is an initiative to virtualize the network services that are now being (http://searchsdn.techtarget.com/definition/network-functions-virtualization-NFV)

Virtual Switch (OVS for example)

- Virtual switches such as Open vSwitch (OVS) are used as the forwarding plane in the hypervisor
- Virtual switches implement extensive support for SDN (e.g. enforce policies) and are widely used by the industry
- Supports L2-L3 networking features:
 - L2 & L3 Forwarding, NAT, ACL, Connection Tracking etc.
 - Flow based
- vSwitch Challenges
 - Awful Packet Performance: <1M w/ 2-4 cores,</p>
 - Burns CPU like Hell : Even w/ 12 cores, can't get 1/3rd 100G NIC Speed
 - Bad User Experience: High and unpredictable latency, packet drops
 - Standard COTS performance is 10x-100x lower than required







Step 1 - Single Root I/O Virtualization (SR-IOV)

- PCIe device presents multiple instances to the OS/Hypervisor
- Enables Application Direct
 Access
 - Bare metal performance for VM
 - Reduces CPU overhead
- Enables many advanced NIC features (e.g. DPDK, RDMA, ASAP^{2,)}





Tradeoff between Virtual Switch and SR-IOV







© 2019 Mellanox Technologies



The Solution - Hardware Acceleration

Software defined everything is a key for modern data centers With today's software based solutions, functionality is (almost) there To gain flexibility, performance and cost efficiency, Hardware Acceleration is needed





Software

Software + Hardware Acceleration

High Performance Workloads Can't Deliver Without HW Acceleration



© 2019 Mellanox Technologies

Software Defined Everything Creates Bottlenecks





Core

Software Defined Everything (SDX) Consumes CPU cores for Packet/Flow Processing and Creates Bottlenecks

- Virtualization, Storage, Switching, Routing, Load Balancing
- Security: Consumes CPU cores for Security Processing
 - Layer 4 Firewall, encryption, host introspection
 - Intrusion detection & prevention





© 2019 Mellanox Technologies

q

One way to get there...

SmartNIC: A network interface card (network adapter) that offloads processing tasks that the system CPU would normally handle. Using its own on-board processor, the SmartNIC may be able to perform any combination of encryption/decryption, firewall, TCP/IP and HTTP processing. SmartNICs are ideally suited for high-traffic Web servers. – PC Magazine Encyclopedia

https://www.pcmag.com/encyclopedia/term/51536/smartnic



© 2019 Mellanox Technologies

SmartNIC – Implementation Varieties





Implementation does Not make a NIC a SmartNIC

System on Chip - NIC + CPU Good price-performance C Programmable Processors Easiest programmability © 2019 Mellanox Technologies 11

Just some of the things you can do...

		Foundational	Intelligent	
Capability	Workloads Accelerated	NIC	NIC (iNIC)	S
	Entry Level Virtualization and Data Moven	nent		
TCP/IP Acceleration	Enterprise workloads	√	√	
NIC Virtualization (SR-IOV)	Enterprise workloads	√	√	
Network Virtualization (VXLAN)	Multi-tenant workloads	√	√	
Da	ata Transport Acceleration & Programmable D)ata Plane		
RoCE Acceleration	VM, Storage, Big Data, Al/ML		√	
DPDK	Virtualized Network Functions		√	
Spark Acceleration	Big Data		√	
NVMe-over-Fabrics Storage	Storage		√	
OVS Hardware Acceleration	Efficient, Scalable Virtualized Apps		√	
QoS and ACL Acceleration	Web Servers / Content Distribution Networks		√	
Flow monitoring/reporting	Visibility, Network Packet Broker, IBN		√	
Flow match/action engine	Software Defined Networking		√	
Fully Programmable Data Plane	Network Function Virtualization		√	
ASAP ² Virtual Switching/Routing	NAT, Load Balancing, stateless firewall		√	
	Smart Networking & Virtualization			
Virtual Switch Policy Engine	OVS Control Plane			
Analytics engine	DPI, Network Monitoring and Diagnostics			
Container Acceleration	Various (AI/ML, Big Data, Analytics)			
Smart Cloud Virtualization	Server Disaggregation & Resource Sharing			
Secur	ity, Compression, Network Function Virtualiza	ation, Storage		
Public Key Crypto, RNG	Authentication, Key Exchange			
Fault Domain Isolation/HA	Bare Metal Cloud			
Stateful IP/ACL filtering	Load Balancing IPD/IDS/UTM			
Storage: Hashing, ECC, Compression	Erasure Coding, Thin Provisioning, Dedup			
Encryption/Decryption	Data at Rest or on-the-fly			
Security VNF Offloading	Firewall, IDS, IPS, Anti-malware, Anti-DDoS			





Accelerated Switching And Packet Processing

A framework that allow offloading Host and/or Application network Data Plane into a NIC

- OpenFlow like (lower granularity)
- Vendor agnostic APIs
 - Critical for an open source solution
 - Community agreed on Linux TC Flower and DPDK RTE-Flow as the main APIs







© 2019 Mellanox Technologies

ASAP²: Full OVS Offload

Enable SR-IOV data path with OVS control plane

In other words, enable support for most SDN controllers with SR-IOV data plane

Use Open vSwitch to be the management interface and offload OVS dataplane to embedded Switch (eSwitch) using acceleration library







© 2019 Mellanox Technologies

ASAP² Direct: Full OVS Offload

Enable SR-IOV data path with OVS control plane

- In other words, enable support for most SDN controllers with SR-IOV data plane
- Use Open vSwitch to be the management interface and offload OVS dataplane to embedded Switch (eSwitch) using acceleration library







© 2019 Mellanox Technologies

OVS over DPDK VS. OVS Offload

Test	OVS Offload	OVS DPDK	Benefit
1 Flow VXLAN	66M PPS	7.6M PPS (VLAN)	8.6X
60K flows VXLAN	19.8M PPS	1.9M PPS	10.4X



OVS Offload provides significant performance boost

Without adding CPU resources





66 MPPS	

© 2019 Mellanox Technologies

Containers too - K8s OVS Offload For CNFs



Similar arch to virtualized environment





© 2019 Mellanox Technologies

OK, so if the SmartNIC is a Computer...





© 2019 Mellanox Technologies

SmartNIC – an Offload Strategy

Commodity NIC



SmartNIC

- Modern data centers suffer from task overload
 - Faster networking, cloud, overlay, SDN & NFV
- Host CPUs end up spending cores on network steering

- SmartNIC frees up host CPUs by utilizing:
 - Hardware accelerations
 - Multicore array of Arm cores
 - Offload the Networking, Storage and Security controls





© 2019 Mellanox Technologies

SmartNIC – Separated Trust Domain Co-Processor







© 2019 Mellanox Technologies

Local Physical Storage to Hardware Emulated Storage





Physical Local Storage

- ✓ Serving bare-metal and hypervisor/VMs
- **X** Bound by physical SSDs capacity
- Vunder-utilized storage
- X Scalability
- X Over-provisioning bound to compute node

Host Server OS/Hypervisor NVMe Standard Driver NVMe Virtual Storage

PCle

BUS

Drive Emulation

- ✓ Serving bare-metal and hypervisor/VMs
- Saving OPEX and CAPEX
- ✓ Supports all network transport types NVMe-oF, iSCSI, iSER and even proprietary







Over-provisioning, scaled to rack/cluster

✓ OS-agnostic using inbox standard NVMe driver

© 2019 Mellanox Technologies

The Path to Accelerated Virtualization



Virtualized Cloud

Freeing up CPU resources

Improve Security & Isolation

Virtualizing Everything



Bare-Metal Cloud

© 2019 Mellanox Technologies





© 2019 Mellanox Technologies

Deploying OpenStack Services at the Network Edge

Common deployment requires 3 servers to run Openstack Control Services

By running Control Services on a SmartNIC, we save 3 servers on Edge site deployment





© 2019 Mellanox Technologies

The next few years...?





If you are interested – the Ethernet NIC Landscape







© 2019 Mellanox Technologies



