SmartNIC:
Accelerating Azure's Network with FPGAs on OCS servers

Daniel Firestone
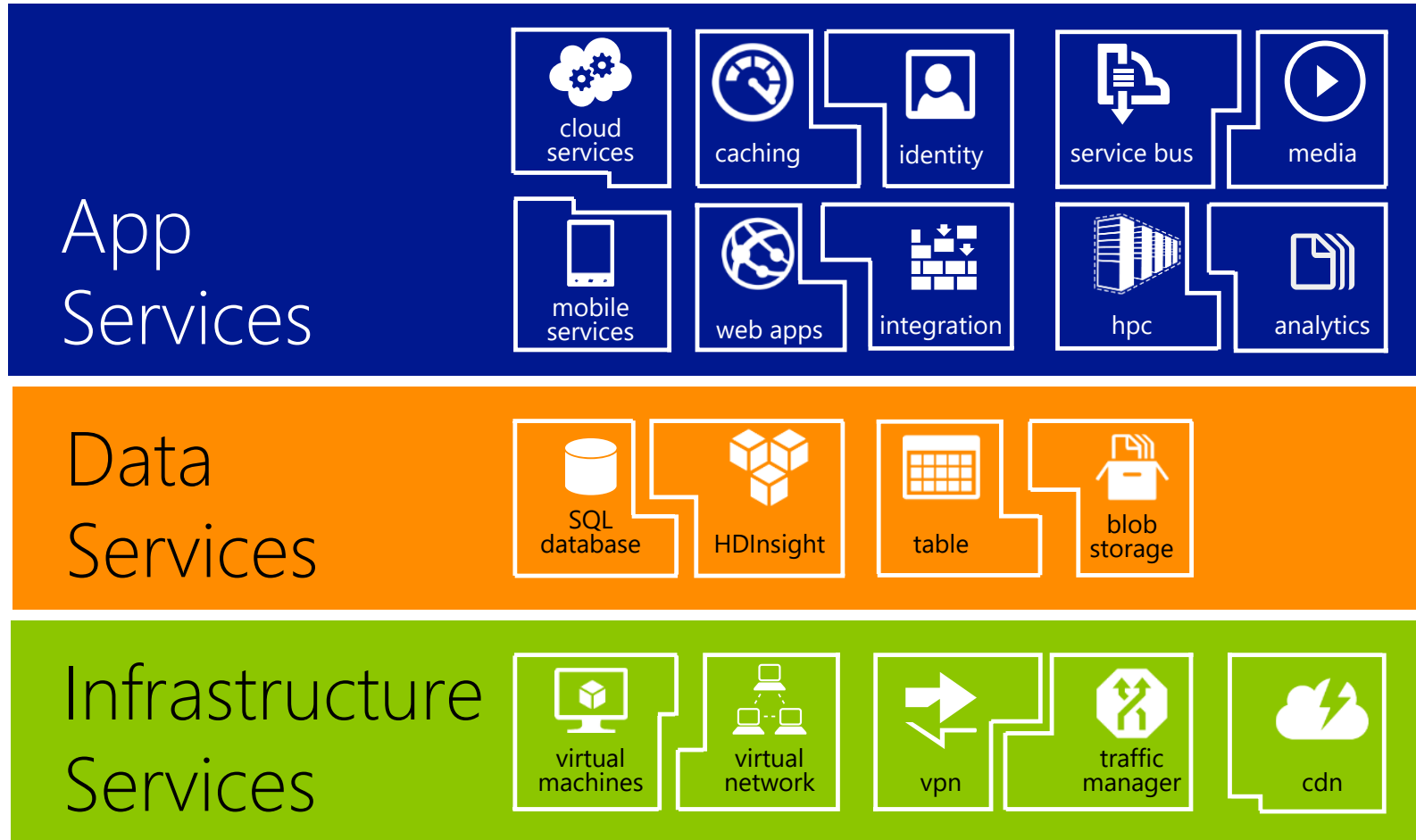
Principal Tech Lead and Software Development Manager

Azure Networking Datapath Team

Microsoft Azure

Summary

- Azure Scale

- Cloud Networking Today: Agility with Software Defined Networking

- Hardware acceleration needed in the 40G+ era

- The industry has relied on ASICs, but ASICs aren't agile enough
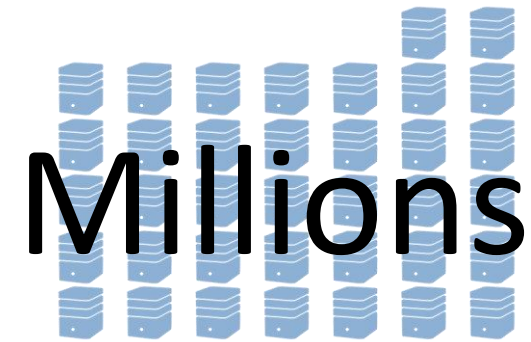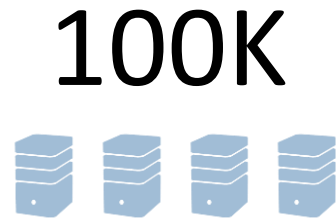
- Solution: FPGA-based SmartNIC
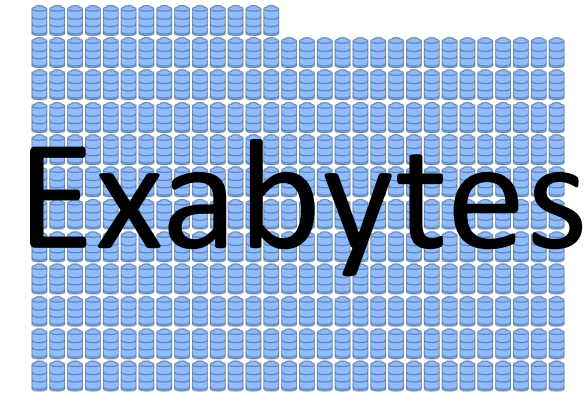
- Demo!

Microsoft Azure

Microsoft Azure

App Services

- cloud services
- caching
- identity
- service bus
- media
- mobile services
- web apps
- integration
- hpc
- analytics

Data Services

- SQL database
- HDInsight
- table
- blob storage

Infrastructure Services

- virtual machines
- virtual network
- vpn
- traffic manager
- cdn

Microsoft Azure

Microsoft Azure

2014
2015
Coming Soon…
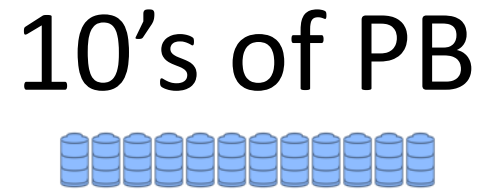
	2010	2016
Compute Instances	100K	Millions
Azure Storage	10's of PB	Exabytes
Datacenter Network	10's of Tbps	Pbps
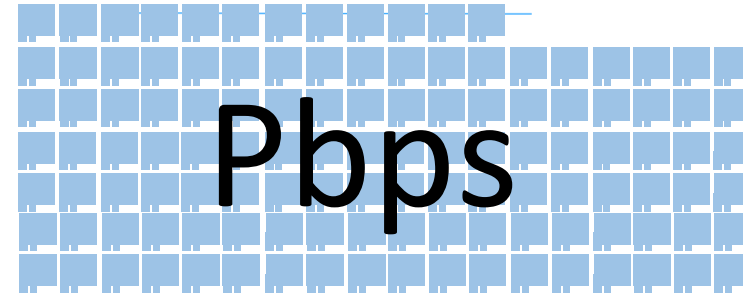
How Do We Build Software Networks in the Cloud?

Microsoft Azure

SDN: Building the right abstractions for Scale

Abstract by separating management, control, and data planes

Example: ACLs

Management plane	Create a tenant
Control plane	Plumb these tenant ACLs to these switches
Data plane	Apply these ACLs to these flows

Data plane needs to apply per-flow policy to millions of VMs
How do we apply billions of flow policy actions to packets?

REST APIs

Management Plane

Controller

Control Plane
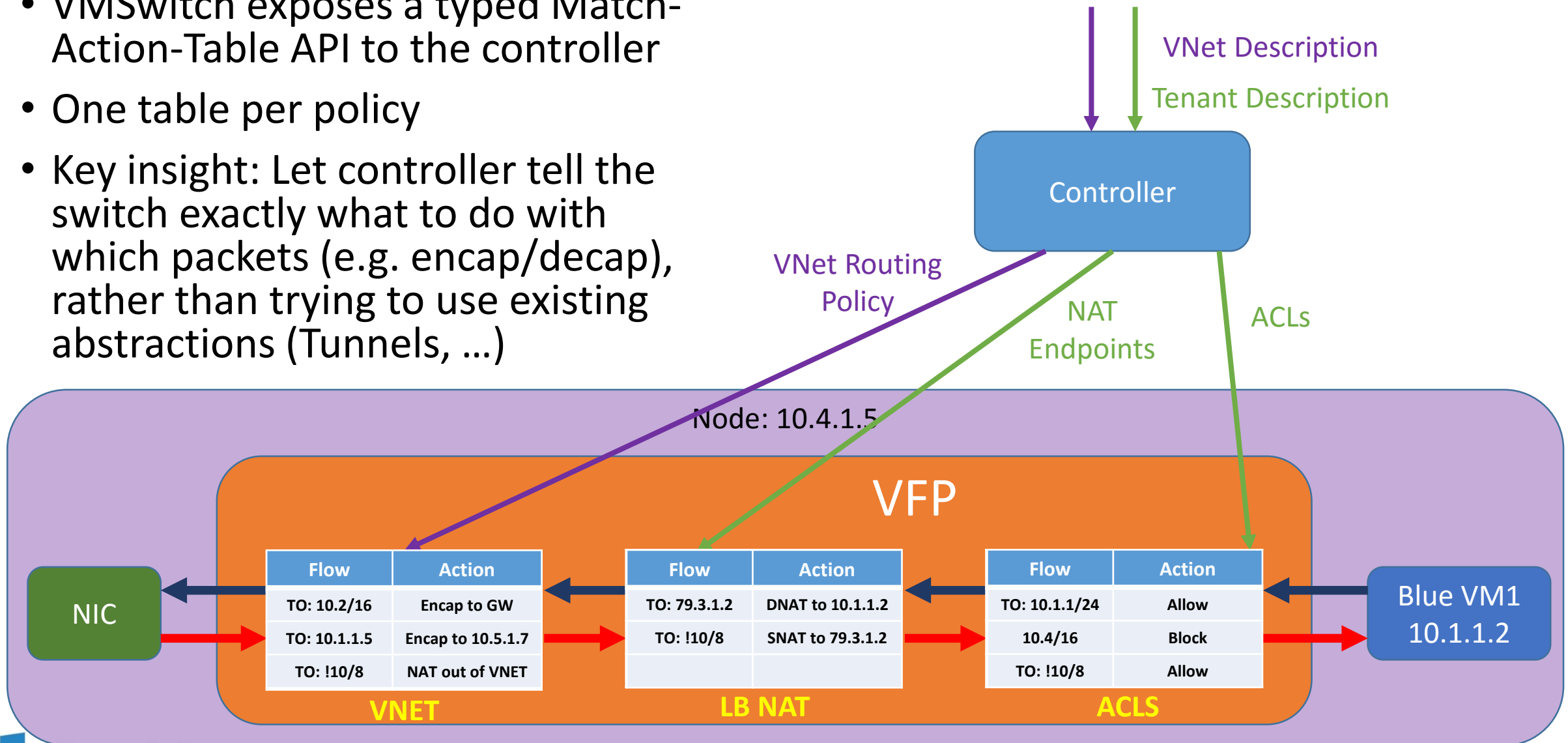
Virtual Switch

Microsoft Azure
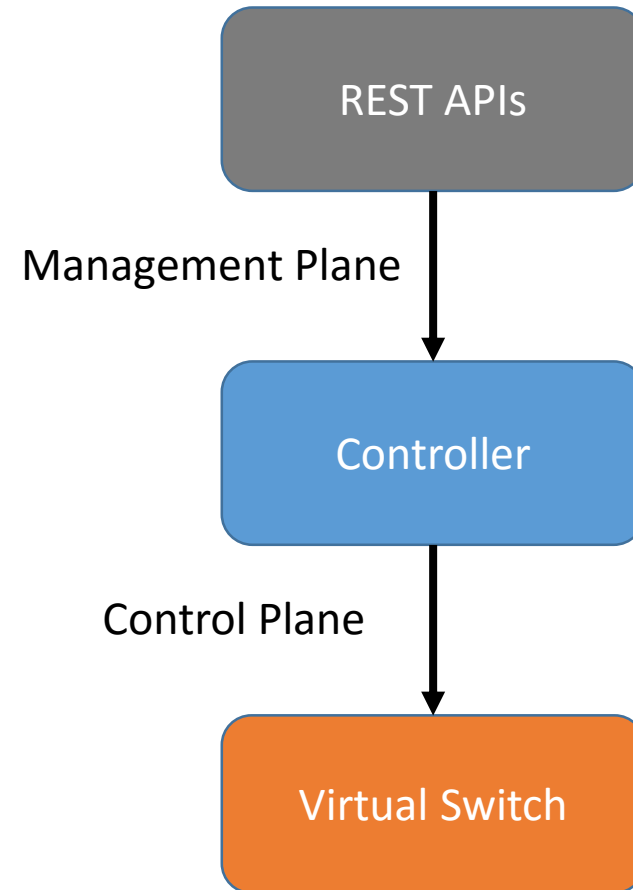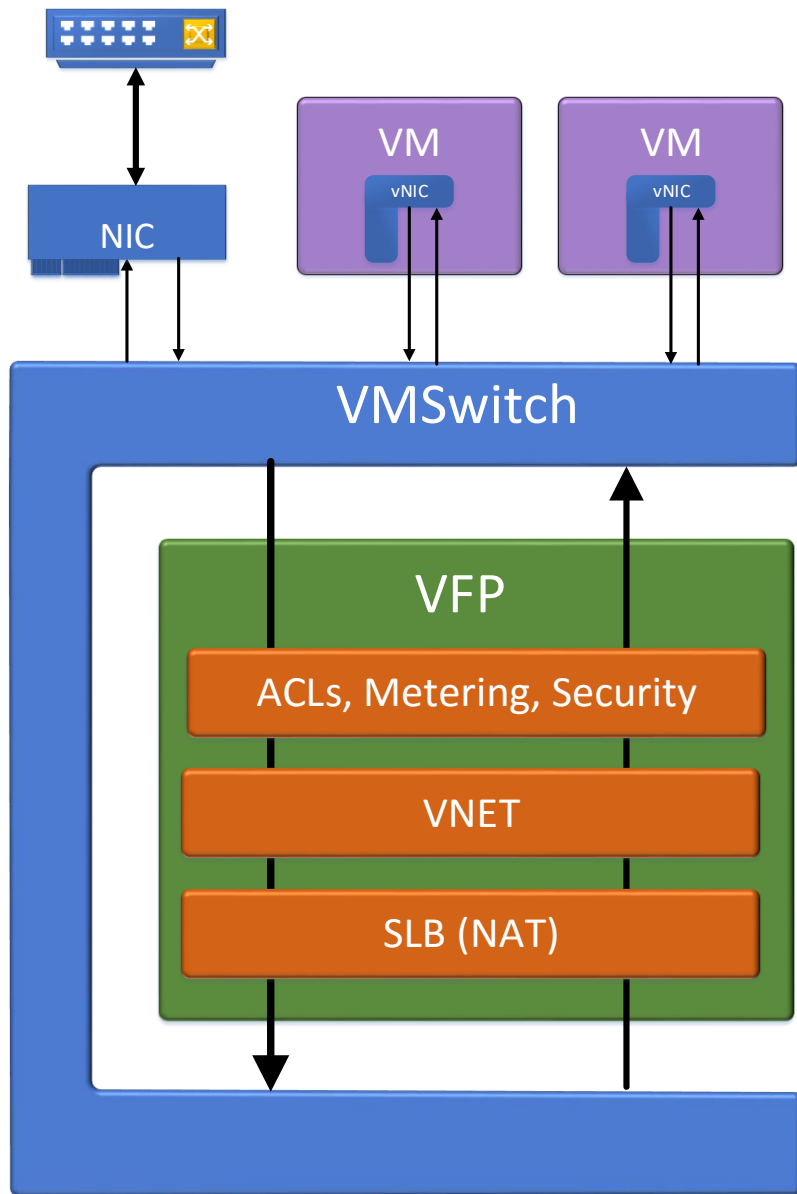
Virtual Filtering Platform (VFP) Azure's SDN Dataplane

- Acts as a virtual switch inside Hyper-V VMSwitch
- Provides core SDN functionality for Azure networking services, including:
  - Address Virtualization for VNET
  - VIP -> DIP Translation for SLB
  - ACLs, Metering, and Security Guards
- Uses programmable rule/flow tables to perform per-packet actions
- Supports all Azure dataplane policy at 40GbE+ with offloads

Microsoft Azure

# Flow Tables are the right abstraction for the Host

- VMSwitch exposes a typed Match-Action-Table API to the controller

- One table per policy

- Key insight: Let controller tell the switch exactly what to do with which packets (e.g. encap/decap), rather than trying to use existing abstractions (Tunnels, …)

VNet Description

Tenant Description

Controller

VNet Routing Policy

NAT Endpoints

ACLs

Node: 10.4.1.5

VFP

NIC

Flow	Action
TO: 10.2/16	Encap to GW
TO: 10.1.1.5	Encap to 10.5.1.7
TO: !10/8	NAT out of VNET

VNET

Flow	Action
TO: 79.3.1.2	DNAT to 10.1.1.2
TO: !10/8	SNAT to 79.3.1.2

LB NAT

Flow	Action
TO: 10.1.1/24	Allow
10.4/16	Block
TO: !10/8	Allow

ACLS

Blue VM1 10.1.1.2

Microsoft Azure

This worked well at 1GbE, ok at 10GbE… what about 40GbE+?

Microsoft Azure

Traditional Approach to Scale: ASICs

- We've worked with network ASIC vendors over the years to accelerate many functions, including:
  - TCP offloads: Segmentation, checksum, …
  - Steering: VMQ, RSS, …
  - Encapsulation: NVGRE, VXLAN, …
  - Direct NIC Access: DPDK, PacketDirect, …
  - RDMA
- Is this a long term solution?

Microsoft Azure
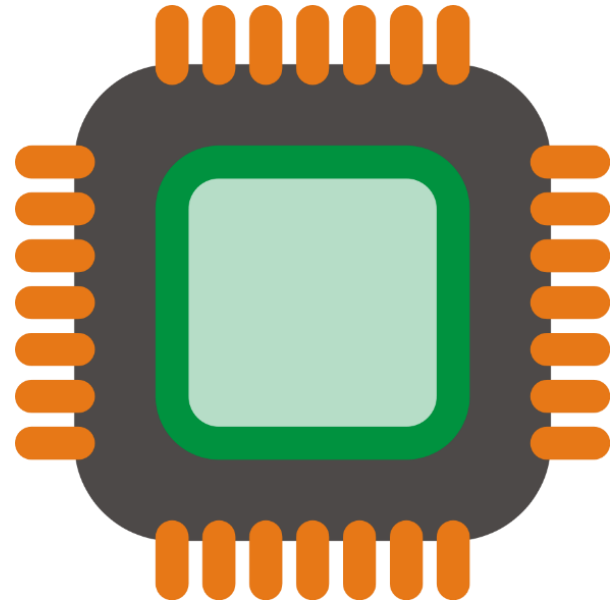
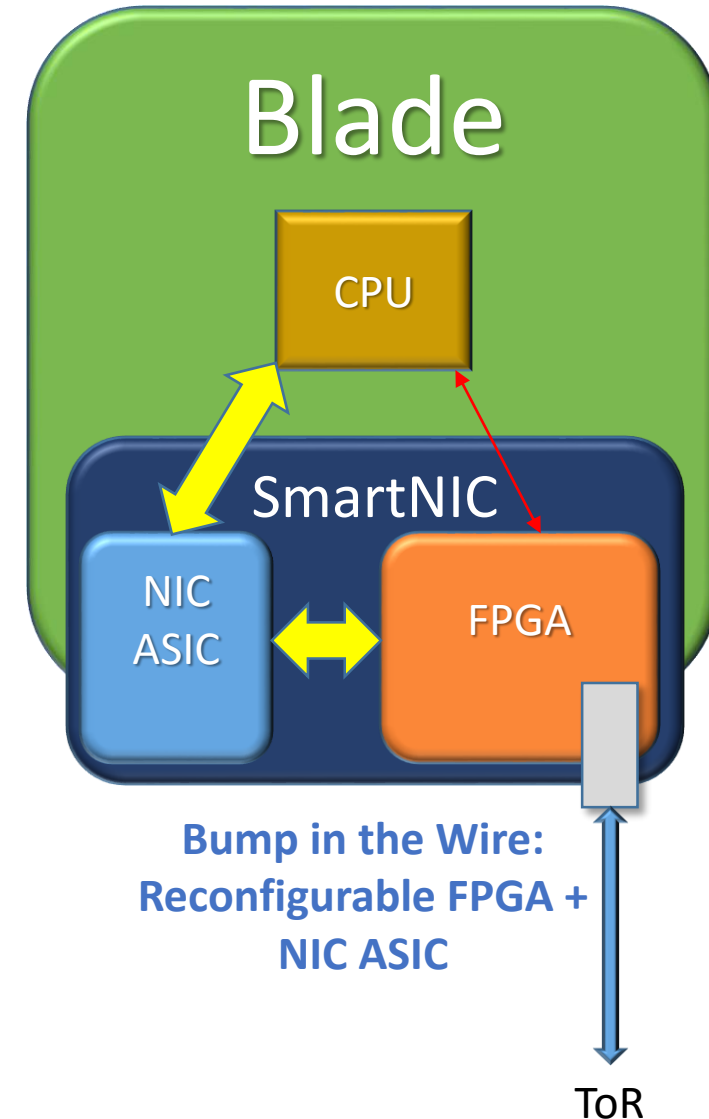# Host SDN Scale Challenges in Practice

- Hosts are Scaling Up: 1G → 10G → 40G → 50G → 100G
  - Reduces COGS of VMs (more VMs per host) and enables new workloads
  - Need the performance of hardware to implement policy without CPU
  - Not enough to just accelerate to ASICs – need to move entire stacks to HW

- Need to support new scenarios: BYO IP, BYO Topology, BYO Appliance
  - We are always pushing richer semantics to virtual networks
  - Need the programmability of software to be agile and future-proof – 12-18 month ASIC cycle + time to roll new HW is too slow

How do we get the performance of hardware

with programmability of software?

Our Solution – Azure SmartNIC

- HW is needed for scale, perf, and COGS at 40G+

- 12-18 month ASIC cycle + time to roll new HW is too slow

- To compete and react to new needs, we need agility – SDN

- SmartNIC combines <u>agility of SDN</u> with <u>speed+COGS of HW</u>

Roll out Hardware as we do Software

Blade

CPU

SmartNIC

NIC ASIC

FPGA

Bump in the Wire: Reconfigurable FPGA + NIC ASIC

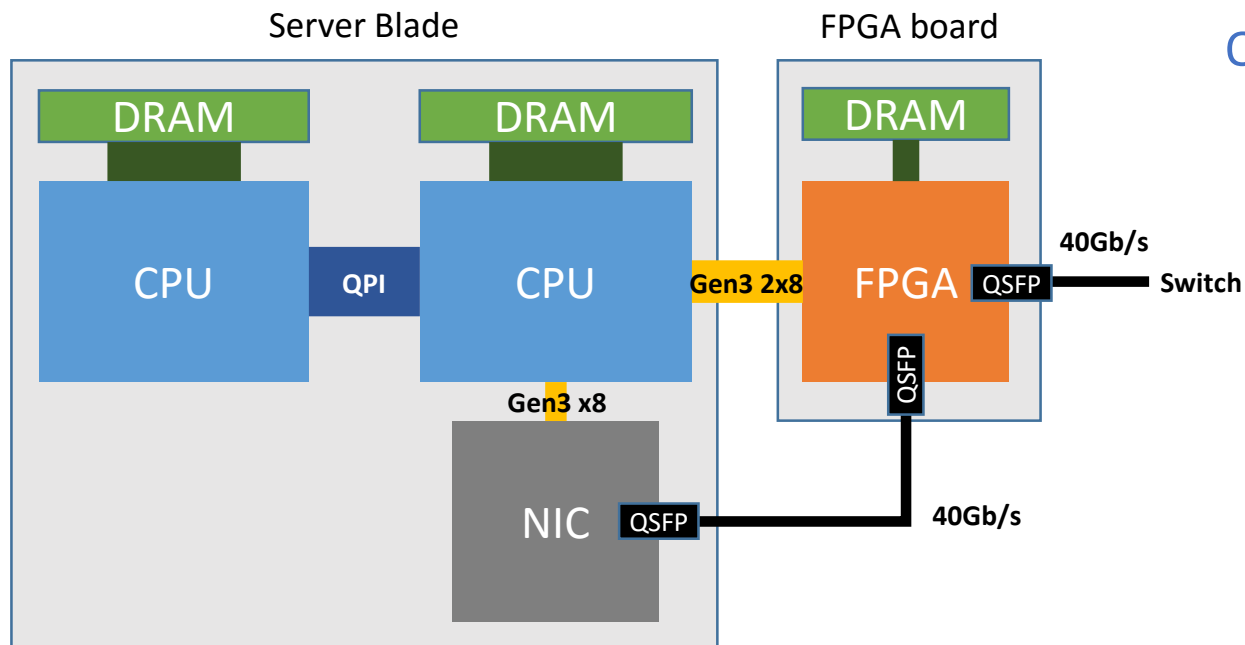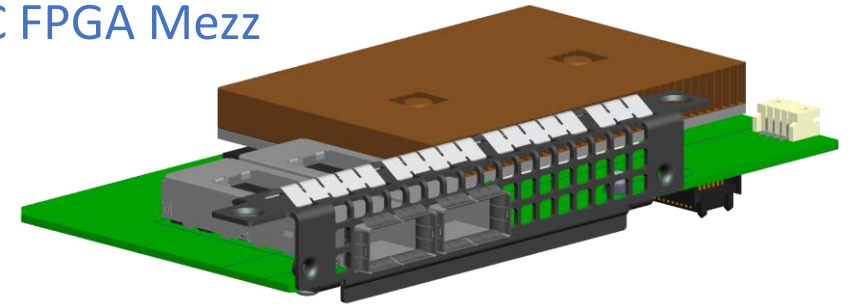ToR

Microsoft Azure

SmartNIC Design

- Use an FPGA for reconfigurable functions
  - FPGAs are already used in Bing
  - Roll out Hardware as we do software
- Programmed using Generic Flow Tables
  - Language for programming SDN to hardware
  - Uses connections and structured actions as primitives
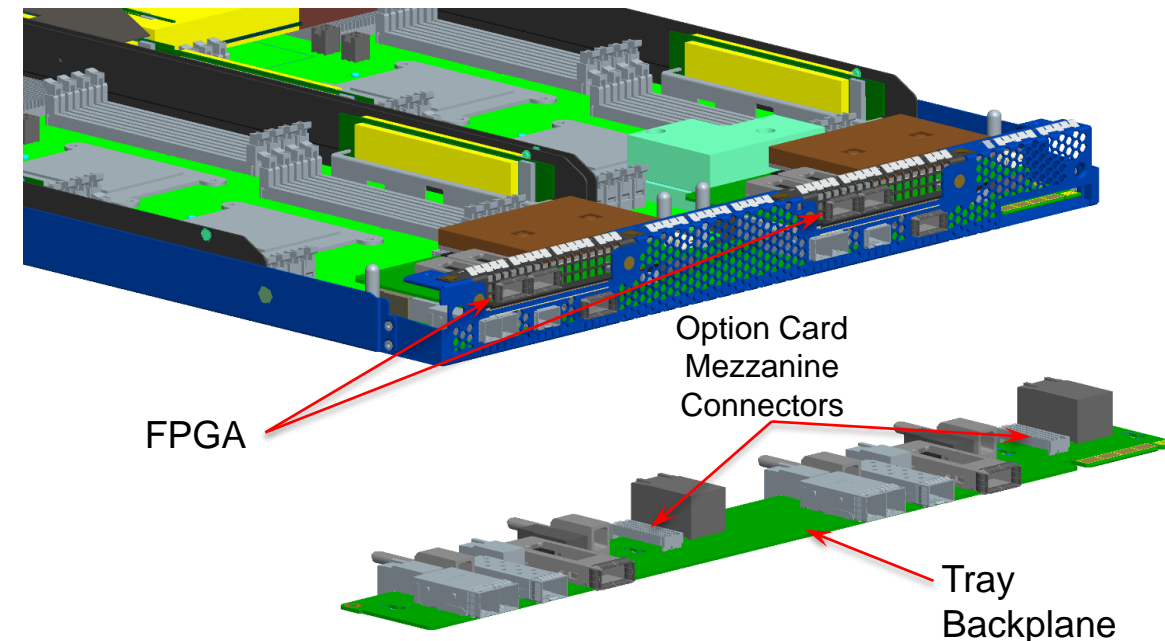- SmartNIC can also do Crypto, QoS, storage acceleration, and more…

Blade

CPU

SmartNIC

NIC ASIC

FPGA

ToR

Microsoft Azure
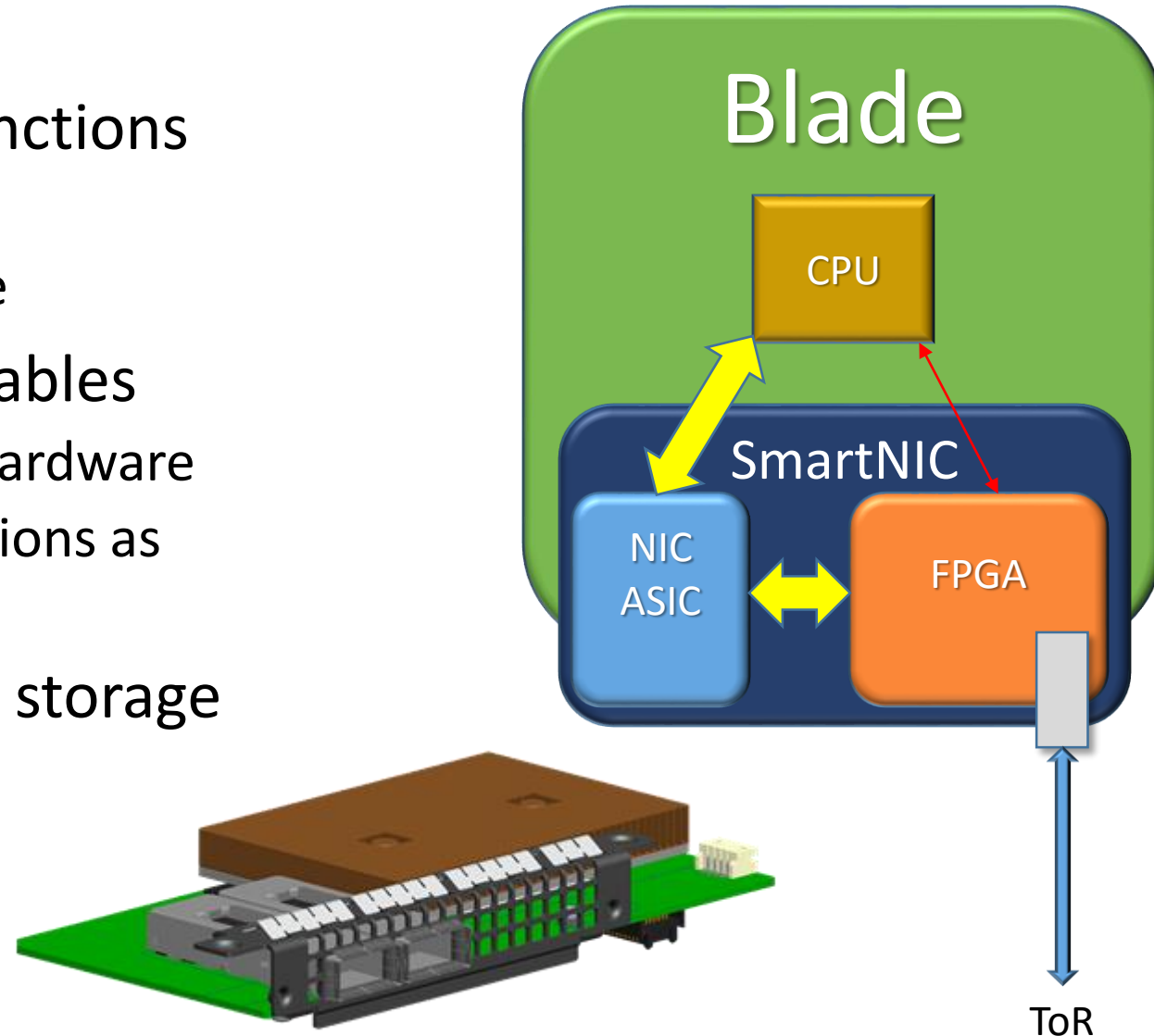
2015 FPGA Deployments: 40G Bump in the Wire

All new Azure Compute servers ship with FPGAs!

SmartNIC FPGA Mezz

Server Blade

FPGA board

DRAM | DRAM | DRAM

CPU | QPI | CPU | Gen3 2x8 | FPGA | QSFP | 40Gb/s | Switch

Gen3 x8

NIC | QSFP | 40Gb/s

OCS Blade with NIC and FPGA

Option Card Mezzanine Connectors

FPGA

Tray Backplane

Microsoft Azure
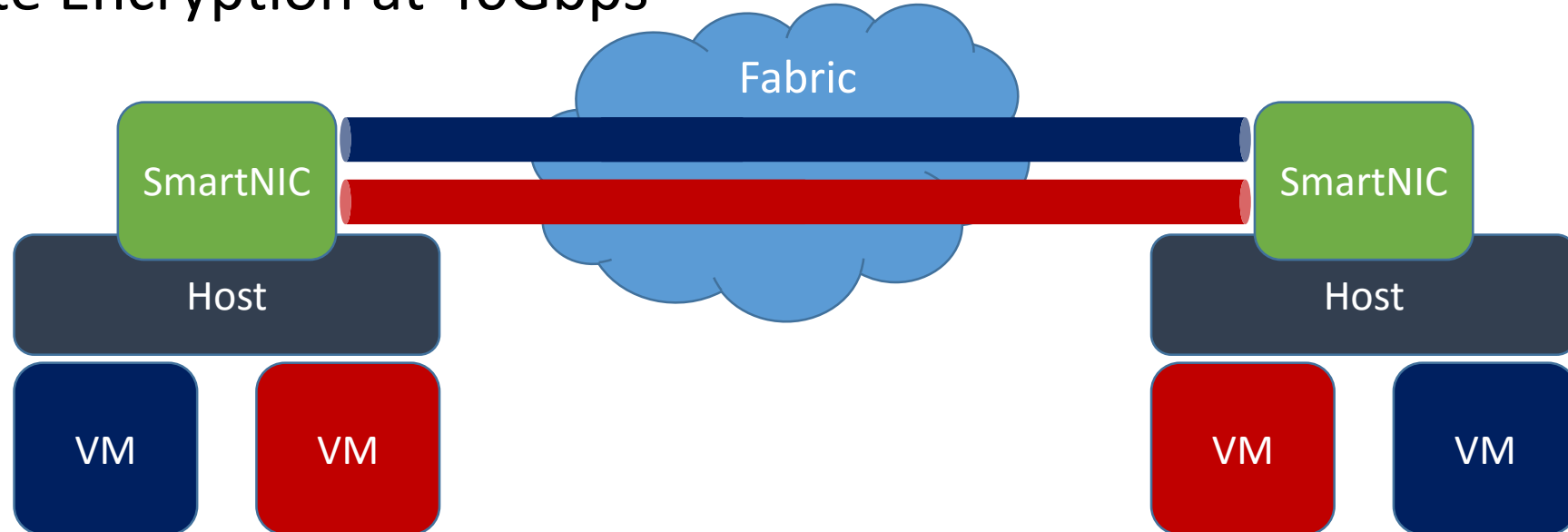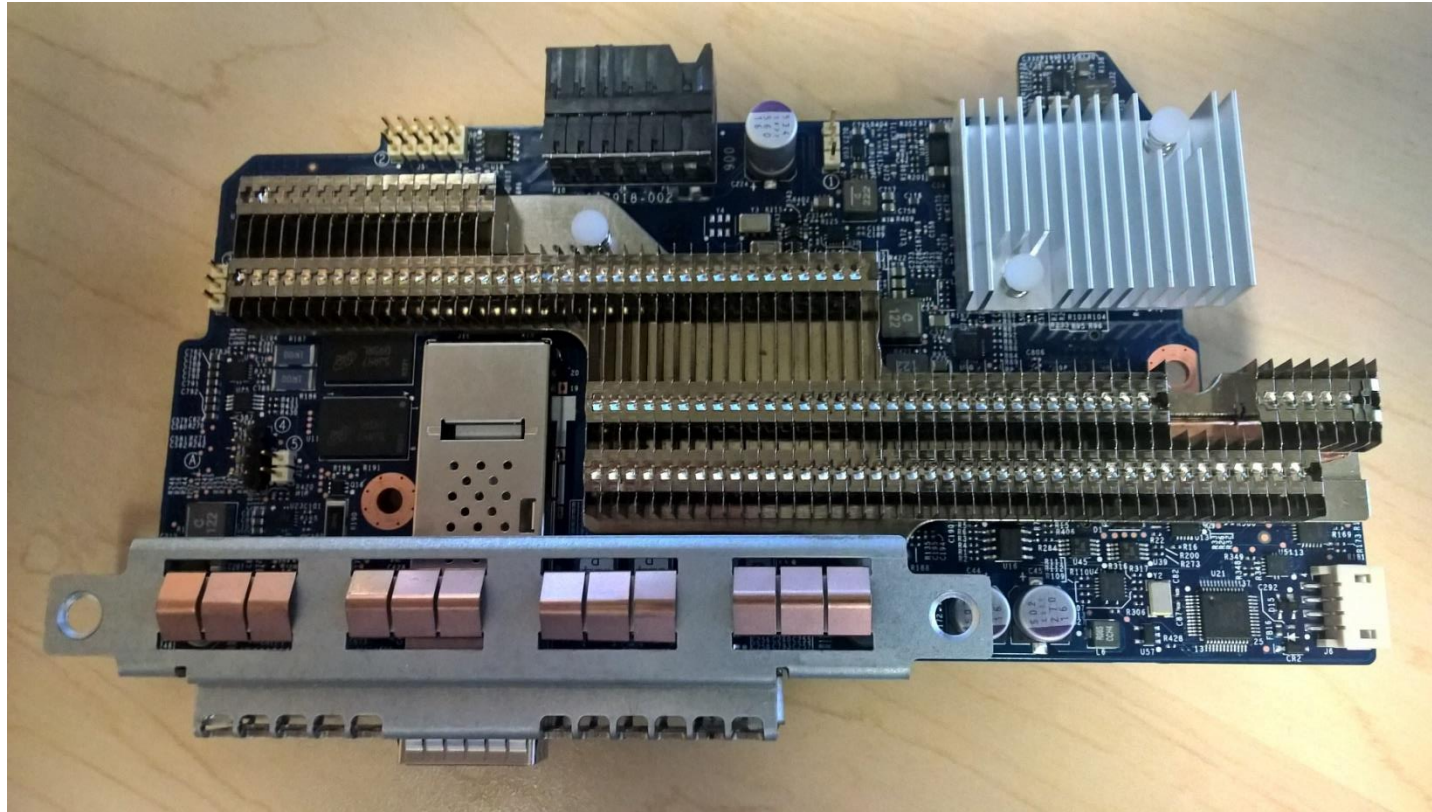
SmartNIC - Accelerating SDN
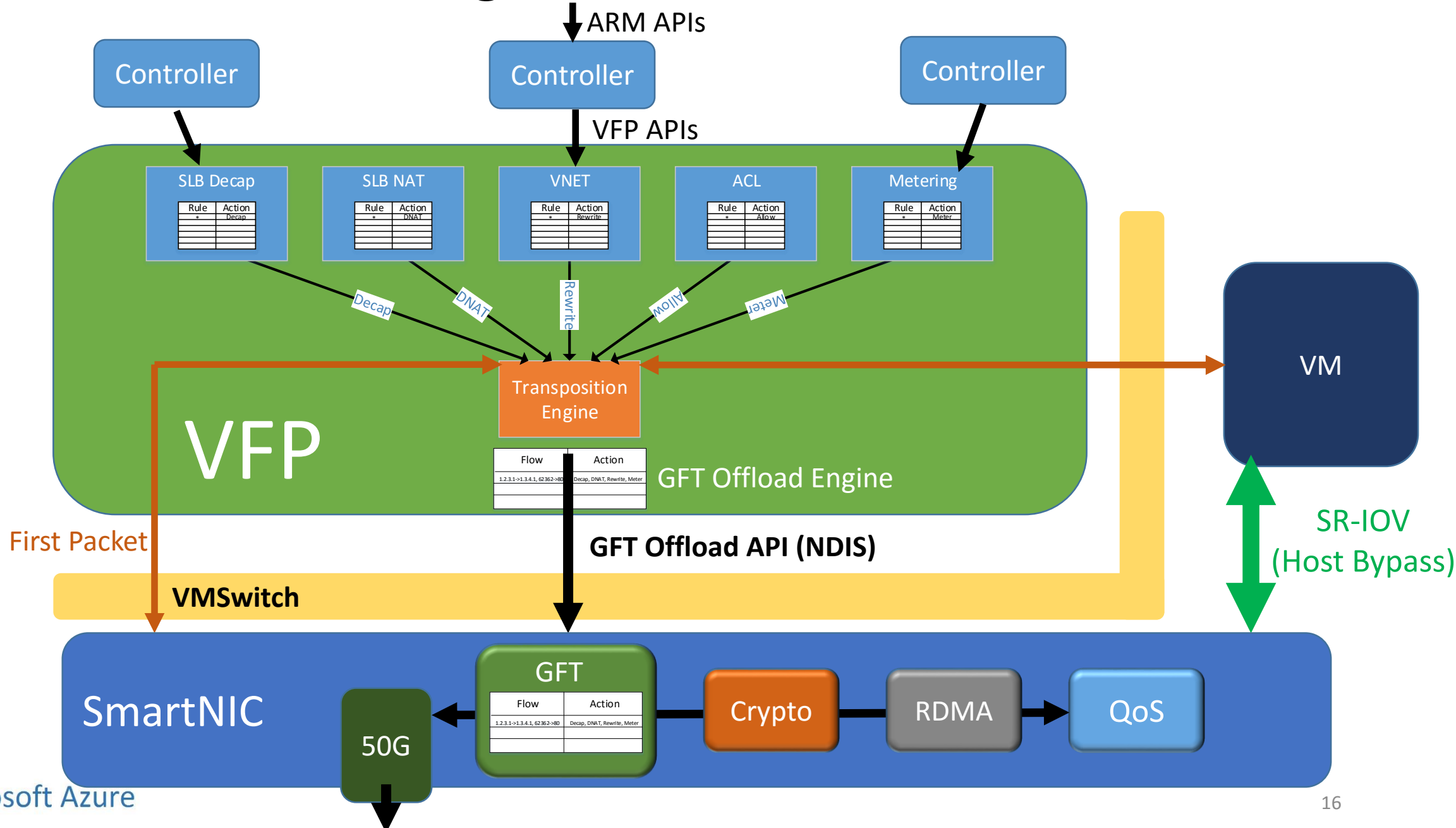
Scenario: Virtual Network Encryption

- SmartNIC can dial encrypted virtual network tunnels (over VxLAN) for each tenant
- Provides E2E security and privacy against actors inside the network fabric
- Line Rate Encryption at 40Gbps

Microsoft Azure

Demo: SmartNIC Encryption

Microsoft Azure

SmartNIC Gen2: Now at 50GbE!

NIC ASIC and FPGA on one Board

Conclusion

- The cloud will continue to scale, and we will continue to add new networking features and scenarios

- ASICs can't keep up with rate of change -> more pressure on FPGAs

- **Ability to change our minds later is the strongest technology we have…**

Want to help lead the reconfigurable computing revolution in the cloud? We're Hiring!

fstone@microsoft.com

Microsoft Azure