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

Daniel Firestone
Principal Tech Lead and Software Development Manager
Azure Networking Datapath Team
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

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
Compute Instances

Azure Storage

Datacenter Network

100K

10’s of PB

10’s of Tbps

2010

Millions

Exabytes

Pbps

2016
How Do We Build Software Networks in the Cloud?
SDN: Building the right abstractions for Scale

Abstract by separating management, control, and data planes

Example: ACLs

<table>
<thead>
<tr>
<th>Management plane</th>
<th>Create a tenant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control plane</td>
<td>Plumb these tenant ACLs to these switches</td>
</tr>
<tr>
<td>Data plane</td>
<td>Apply these ACLs to these flows</td>
</tr>
</tbody>
</table>

Data plane needs to apply per-flow policy to millions of VMs

How do we apply billions of flow policy actions to packets?
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
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, ...)

![Diagram showing flow tables and network operations]
This worked well at 1GbE, ok at 10GbE... what about 40GbE+?
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?
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 agility of SDN with speed+COGS of HW

Roll out Hardware as we do Software
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…
2015 FPGA Deployments: 40G Bump in the Wire

All new Azure Compute servers ship with FPGAs!
SmartNIC - Accelerating SDN

Controller

Controller

Controller

SLB Decap
SLB NAT
VNET
ACL
Metering

Decap, DNAT, Rewrite, Meter
1.2.3.1->1.3.4.1, 62362->80

Decap
DNAT
Rewrite
Allow
Meter

VFP

VMSwitch

SmartNIC

GFT Offload API (NDIS)

GFT Offload Engine

SR-IOV (Host Bypass)

VFP APIs

ARM APIs

VFP

First Packet

SMARTNIC

GFT

Crypto

RDMA

QoS

50G

Microsoft Azure
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
Demo: SmartNIC Encryption
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