Engineering SDN for Scale

Deepak Bansal
Partner Development Manager
Microsoft Azure
dbansa@Microsoft.com
1.4 million miles of fiber in our DCs
>80% of Fortune 500 use the Microsoft Cloud
SDN Motivation and Challenges

- **Enable customers to run the cloud services with similar or better network than on-premise**
  - Per customer network, with richness, flexibility, control, isolation and programmability of on-premise network (firewalls, routes, load balancing, DMZ, VPN etc)
  - Virtual appliances in the “virtual network” in the cloud

- **Deploying and managing complex policies on physical devices does not scale**

- **Challenge:** How do we deliver and scale virtual networks across millions of servers?

- **Solution:**
  - Network Controller: Centralized, scalable, highly available, goal state based management of the customer’s virtual networks
  - Host SDN: scalable virtual network data plane via host agents and host drivers
  - Virtual network functions: Provide virtual network services like load balancing and address resolution implemented as services on Azure
Network Controller

• Challenges:
  • Scale
    • 1DC/1 region -> ~50DCs/17 regions
    • 1000s of virtual networks and network endpoints -> 100,000s -> millions
  • Rate of virtual network provisioning
    • 30min provisioning time -> <5s provisioning time with containers
  • Scope of virtual network
    • Within cluster -> region -> global

• Solution: Scalable and highly available network controller that is
  • Hierarchical
  • Partitioned and regional
  • Micro-services based
Hierarchical Network Controller

Regional

- NRP
- RNM

Cluster

- NSM
- CDS

Node

- NM Agent
Regional Network Controller Stats

• 10s of millions of API calls per day
• API execution time
  • Read : <50 milliseconds
  • Write : <150 milliseconds
• Varying deployment footprint
  • Smallest : <10 Hosts
  • Largest : >100 Hosts
Network Controller

RNM

Controller

Node1: 10.1.1.5
Blue VM1 10.1.1.2
Green VM1 10.1.1.2
Azure VMSwitch

Node2: 10.1.1.6
Red VM1 10.1.1.2
Green VM2 10.1.1.3
Azure VMSwitch

Node3: 10.1.1.7
Green S2S GW 10.1.2.1

Secondary Controllers

Consensus Protocol

Customer Config
Allocated Network Resources
L3 Forwarding Policy

Green Enterprise Network 10.2/16
VPN GW

Customer Config
Allocated Network Resources
L3 Forwarding Policy

Secondary Controllers
Consensus Protocol

Node1: 10.1.1.5
Azure VMSwitch
Green VM1 10.1.1.2
Blue VM1 10.1.1.2

Node2: 10.1.1.6
Azure VMSwitch
Red VM1 10.1.1.2
Green VM2 10.1.1.3

Node3: 10.1.1.7
Azure VMSwitch
Green S2S GW 10.1.2.1
SLB Components:
- SLBM (SLB manager): Control plane that tracksvip dip mappings, health and programs Muxes and hosts
- Mux: Data plane that forwards traffic as well as announce routes
- SLB Host agent: Does NATting

SLB Scale
- 600+ SLBs in production
- 1 Tbps (aggregated) of data
- 17 million of SNAT slot requests
- 1 million total endpoints
Virtual Filtering Platform (VFP)

- 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 data plane policy at 40GbE+ with offloads
- Coming to private cloud in Windows Server 2016
Flow Tables: the Right Abstraction for the Host

VMSwitch exposes a typed Match-Action-Table API to the controller
- Controllers define policy
- One table per policy

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

VNet Description
VNet Routing Policy
NAT Endpoints
ACLS

Controller

NIC
VNET
VFP
LB NAT
ACLS

VM1

Host: 10.4.1.5

Flow | Action
--- | ---
TO: 10.2/16 | Encap to GW
TO: 10.1.1.5 | Encap to 10.5.1.7
TO: 110/8 | NAT out of VNET

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

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

Controller to VNet Description to Tenant Description
VNet Routing Policy to ACLs
We are Hiring!

• BA/BS, MS, PhD
• Interns, Full time positions

{Dbansal, parveenp, umkrishn}@microsoft.com