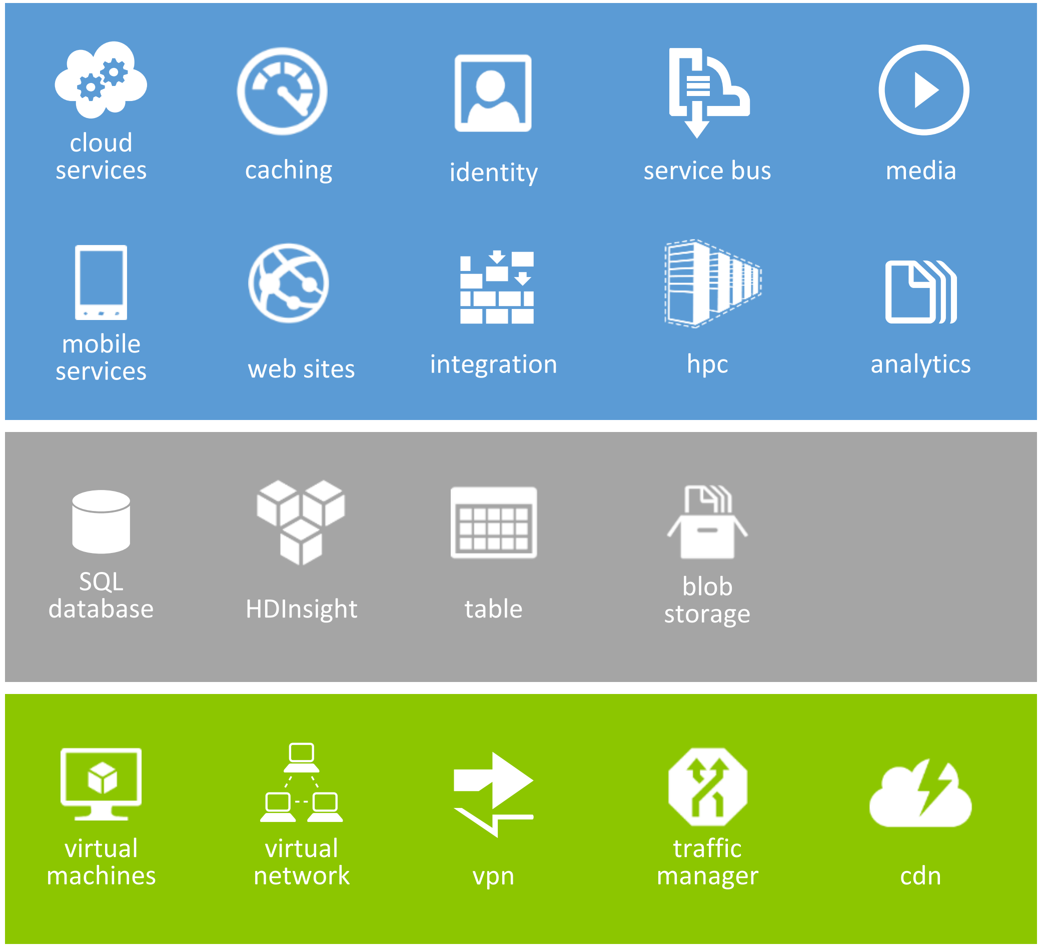
SDN in the Public Cloud: Windows Azure

Albert Greenberg
Partner Development Manager
Windows Azure Networking
albert@microsoft.com

Windows Azure

- Microsoft's big bet on public cloud service
- Lets companies move their IT infrastructure to the cloud
- Provides platform services to build SaaS applications, infrastructure services for IT, scalable cloud storage, and more
- Elastic scaling and much lower COGS than on-prem data centers
- Also runs major Microsoft cloud properties. All are moving to Windows Azure



cloud services

caching

identity

service bus

media

mobile services

web sites

integration

hpc

analytics

SQL database

HDInsight

table

blob storage

virtual machines

virtual network

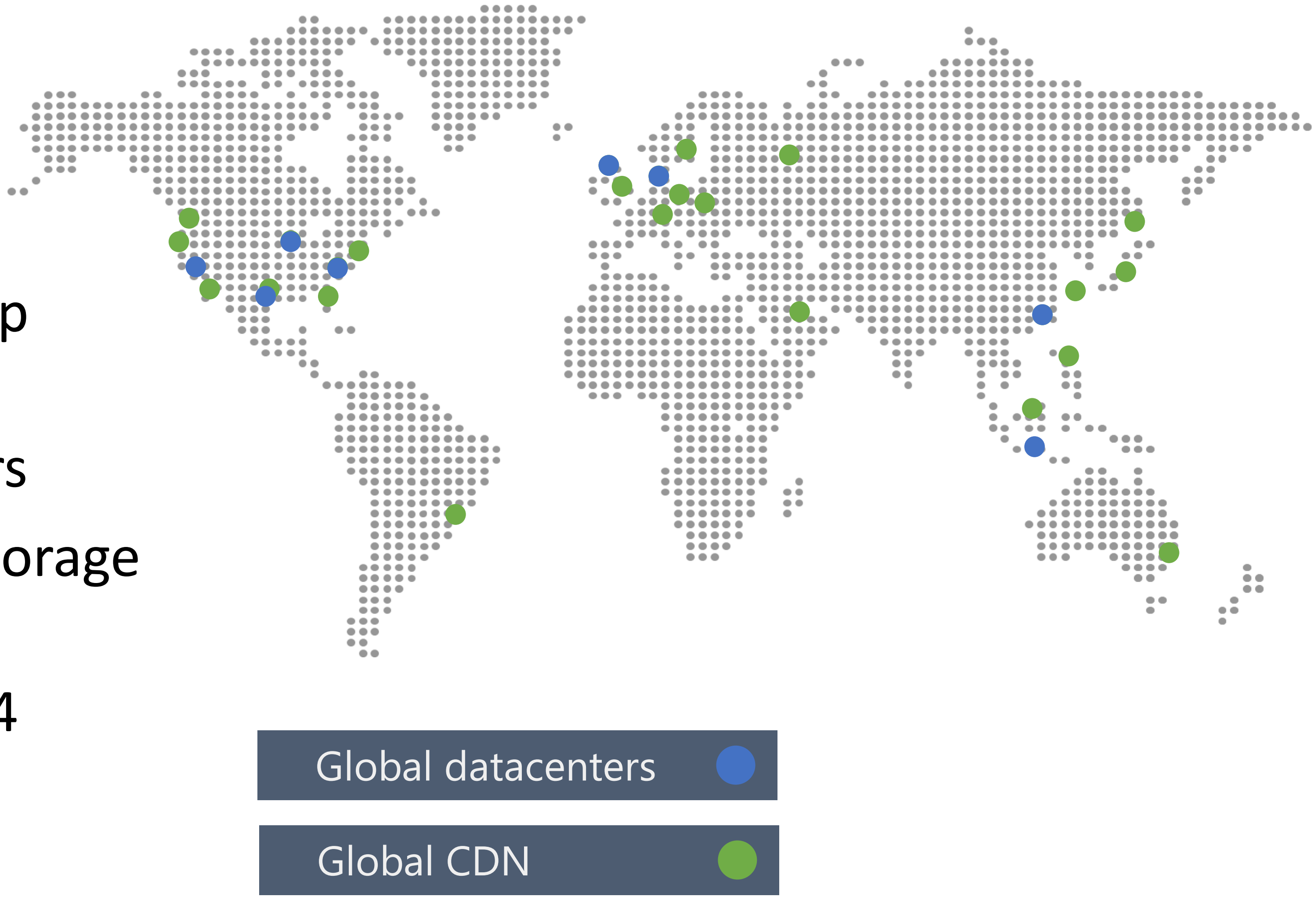
vpn

traffic manager

cdn

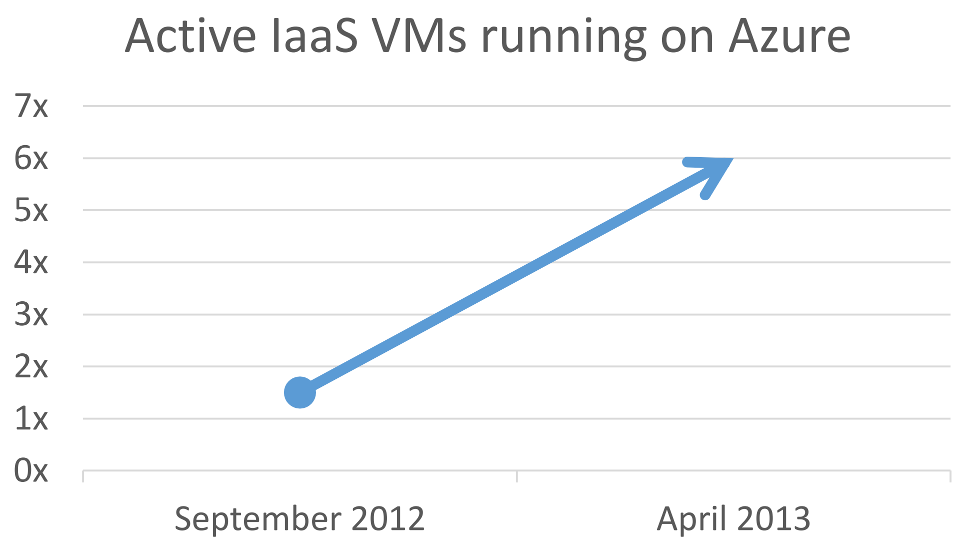
Windows Azure - Some Stats

- More than 50% of Fortune 500 companies using Azure
- Nearly 1000 customers signing up every day
- Hundreds of thousands of servers
- We are doubling compute and storage capacity every 6-9 months
- Azure Storage is Massive – over 4 trillion objects stored
- Windows Azure Directory has processed 200 billion authentications



Big Bet on Enterprise: Infrastructure as a Service (IaaS)

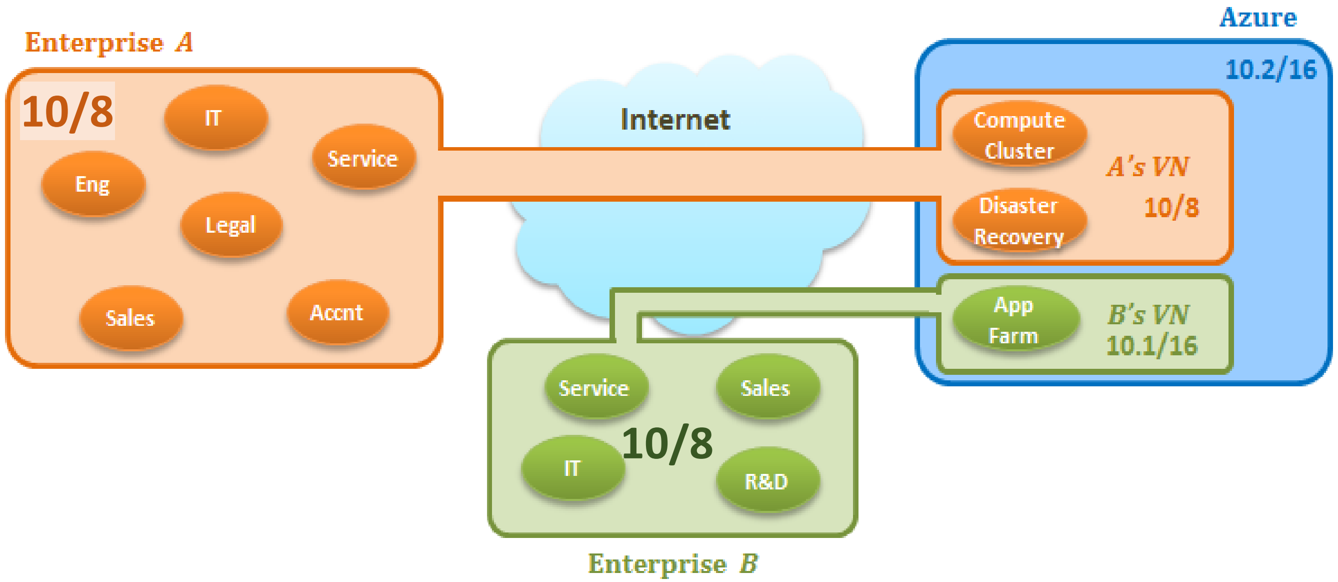
- Bring your own persistent VMs to the Cloud
  - Anything that runs on Hyper-V runs on Azure
- Manage your infrastructure and policy as if it was your own
- What does IaaS require of networking?
  - Domain join your VMs to your domain controller
  - Connect cloud DNS to on-premise DNS
  - Set up a Sharepoint farm in the cloud and serve it back to your corporate intranet



1.5 Million IaaS VMs created since IaaS preview (June ’12)

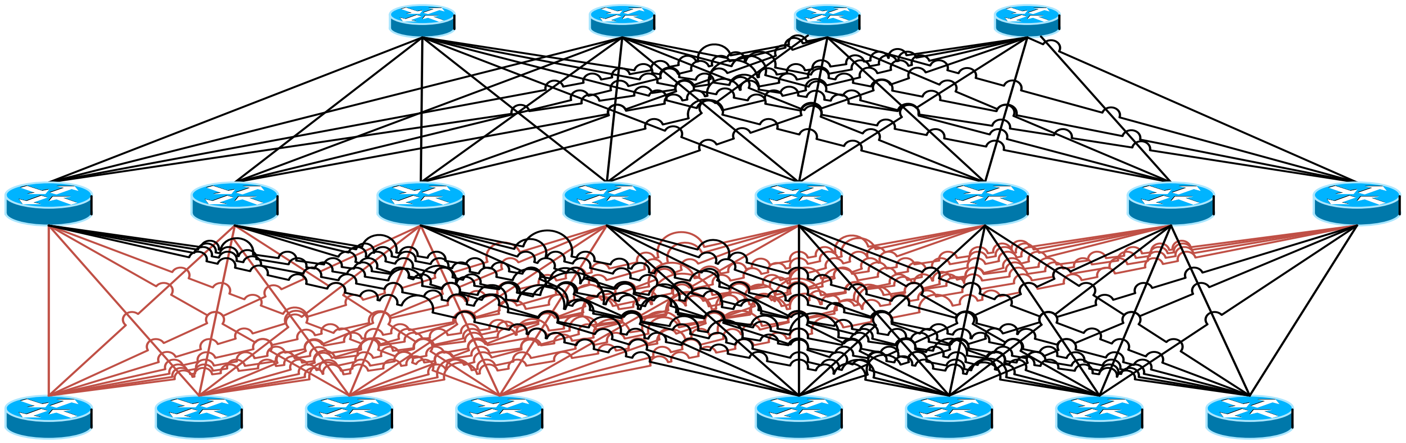
Onboarding Enterprises to the Cloud: Windows Azure Virtual Networks

- Goal: Windows Azure is just another branch office of your enterprise, via VPN
- Requirements
  - BYO Customer Address (CA) Space + Policy to the cloud
  - Communication between tenants of your Azure deployment is efficient and scalable
- Enabled via overlay networking (NVGRE)



Challenge of Building VNet: Agility at Scale

- Goals
  - Agility: On demand provisioning of customer networks and policies
    - Every time a customer creates a VM or tenant, network resources must be provisioned
  - Scale: Millions of VMs, 100's of thousands of nodes with high density (10GbE)
- How to solve this: division of labor
  - **Software focuses on translation of demands to policy and implementation**
  - **Hardware focuses on capacity, reliability and perf**



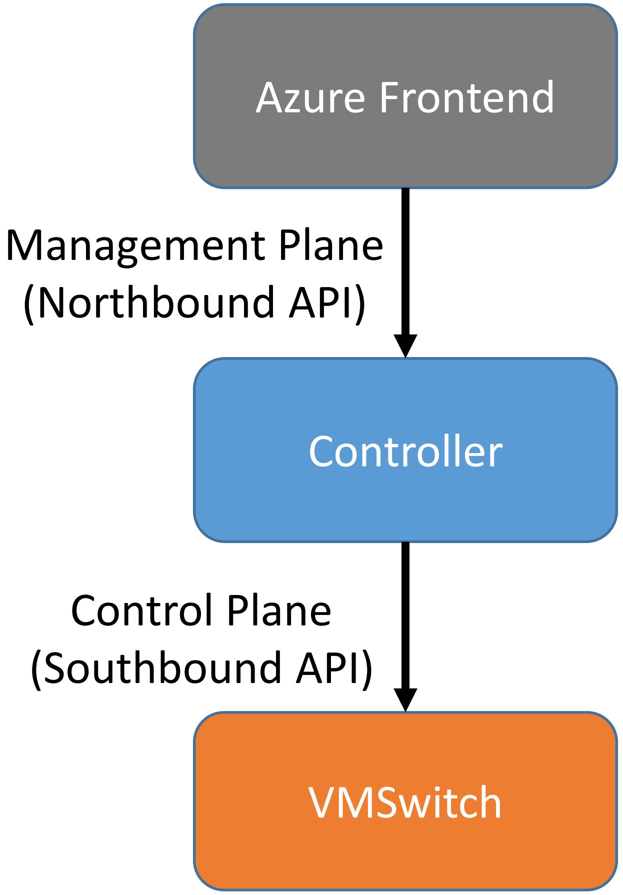
10G Servers

Solution: Software Defined Networking

- Key architecture: policy is abstracted, with sharp division of labor between management, control and data planes

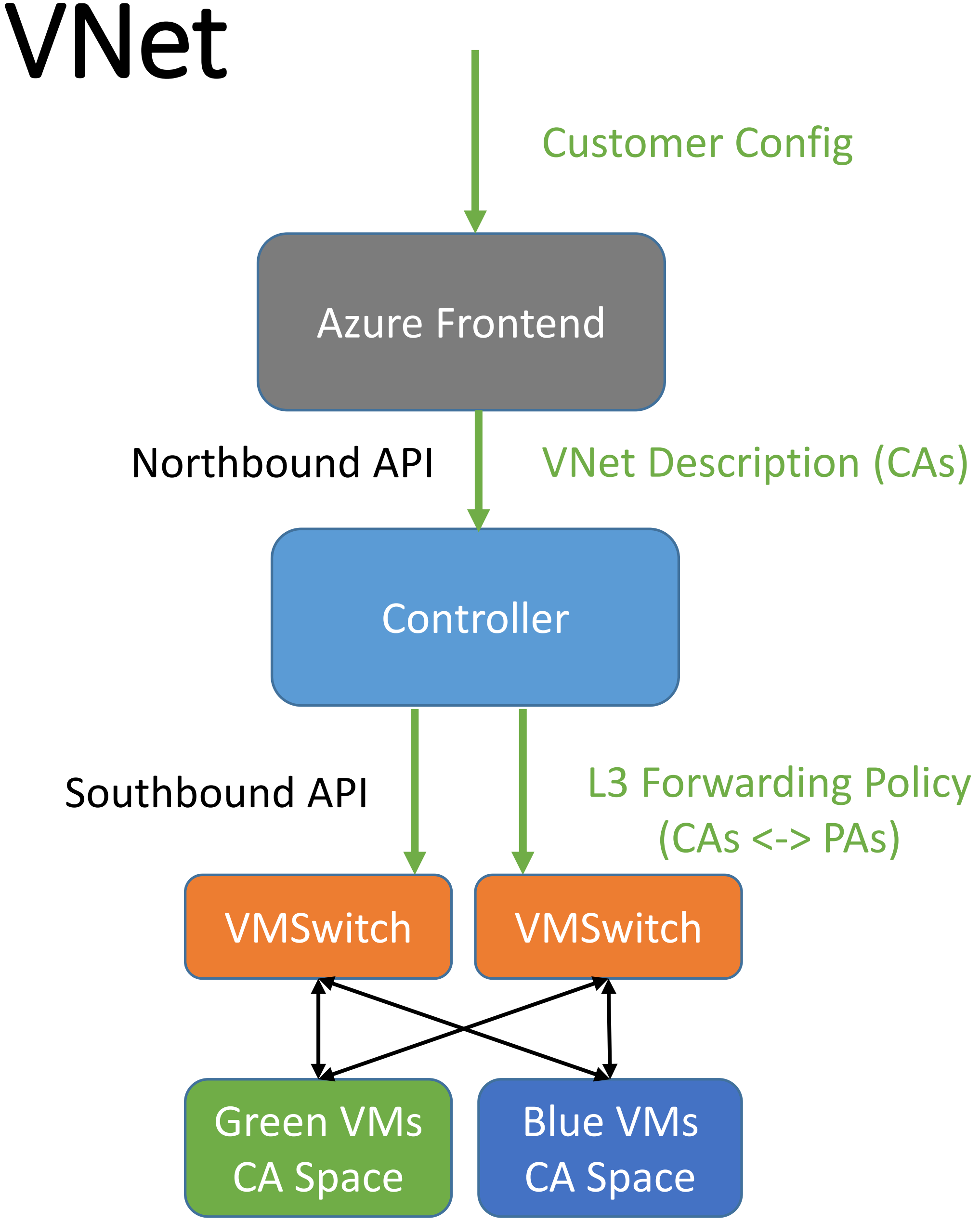
Example: Access Control Lists (ACLs)

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

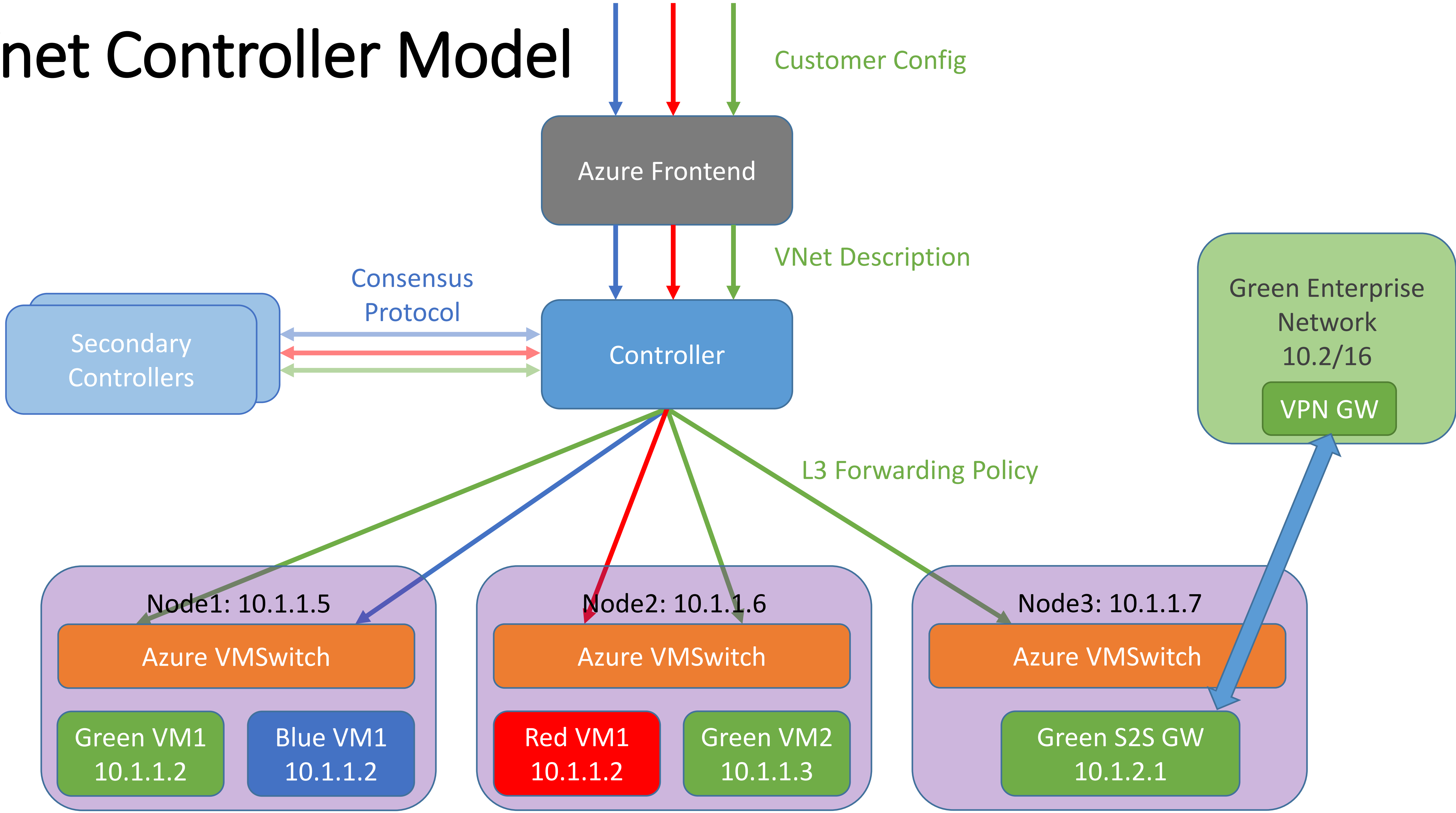


SDN Approach to Building VNet

- A VNet is essentially a set of mappings from a customer defined address space (CAs) to provider addresses (PAs) of hosts where VMs are located
- Separate the interface to specify a VNet from the interface to plumb mappings to VMSwitches via a Network Controller

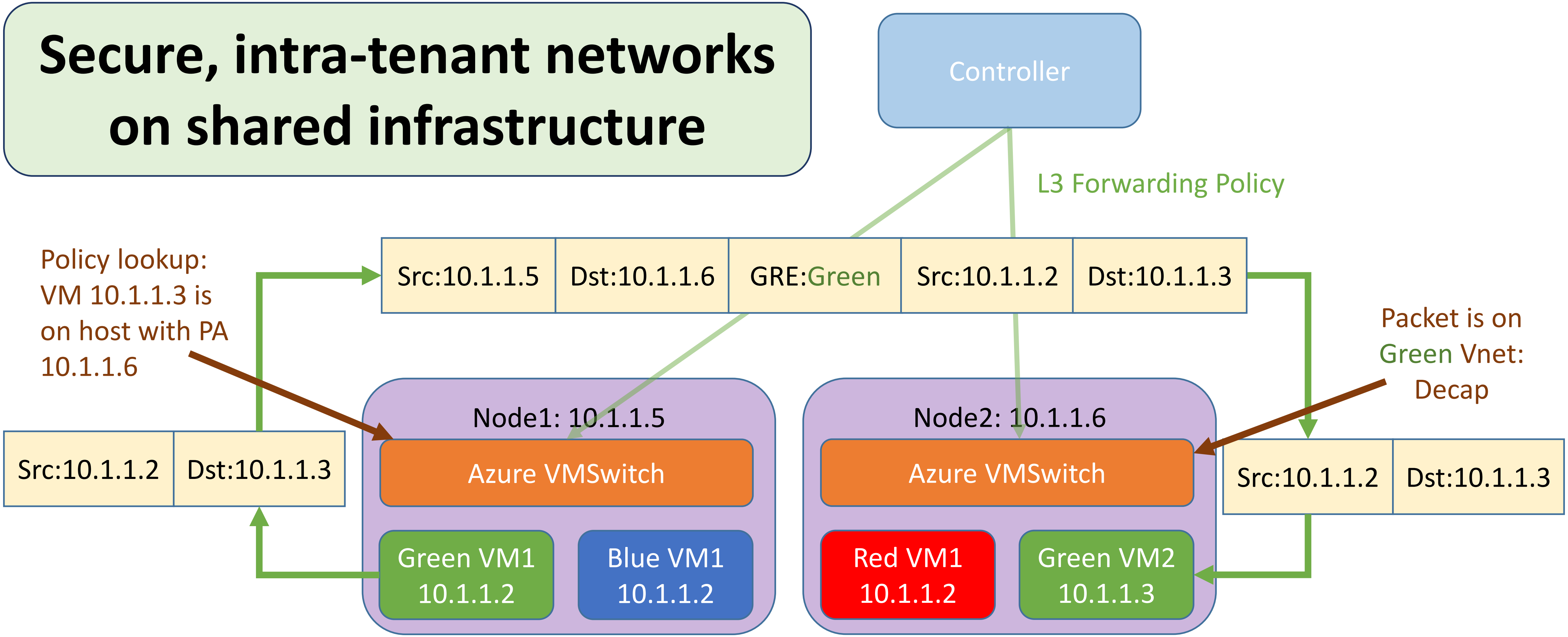


Vnet Controller Model

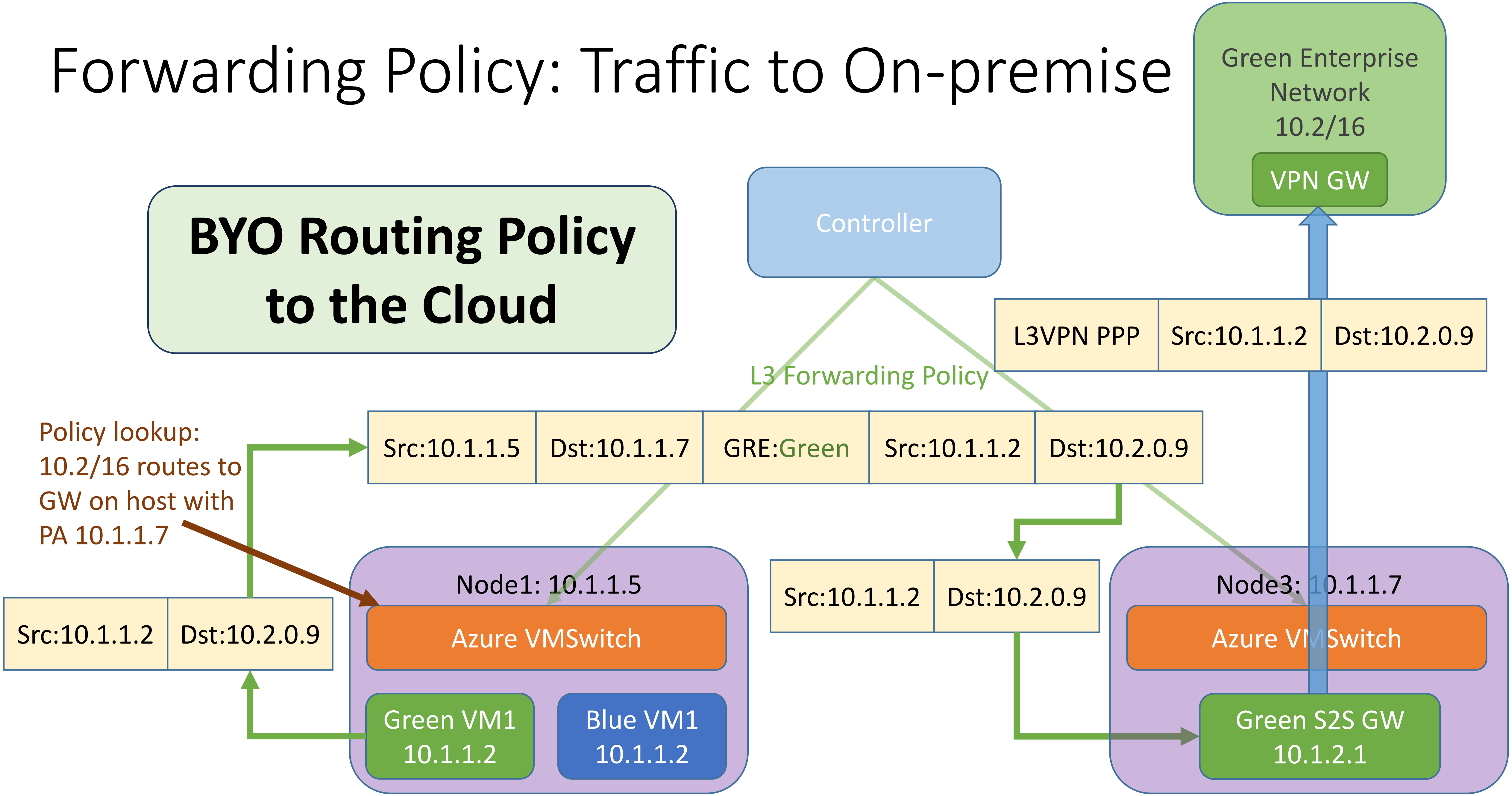


Forwarding Policy: Intra-VNet

Secure, intra-tenant networks on shared infrastructure



Forwarding Policy: Traffic to On-premise

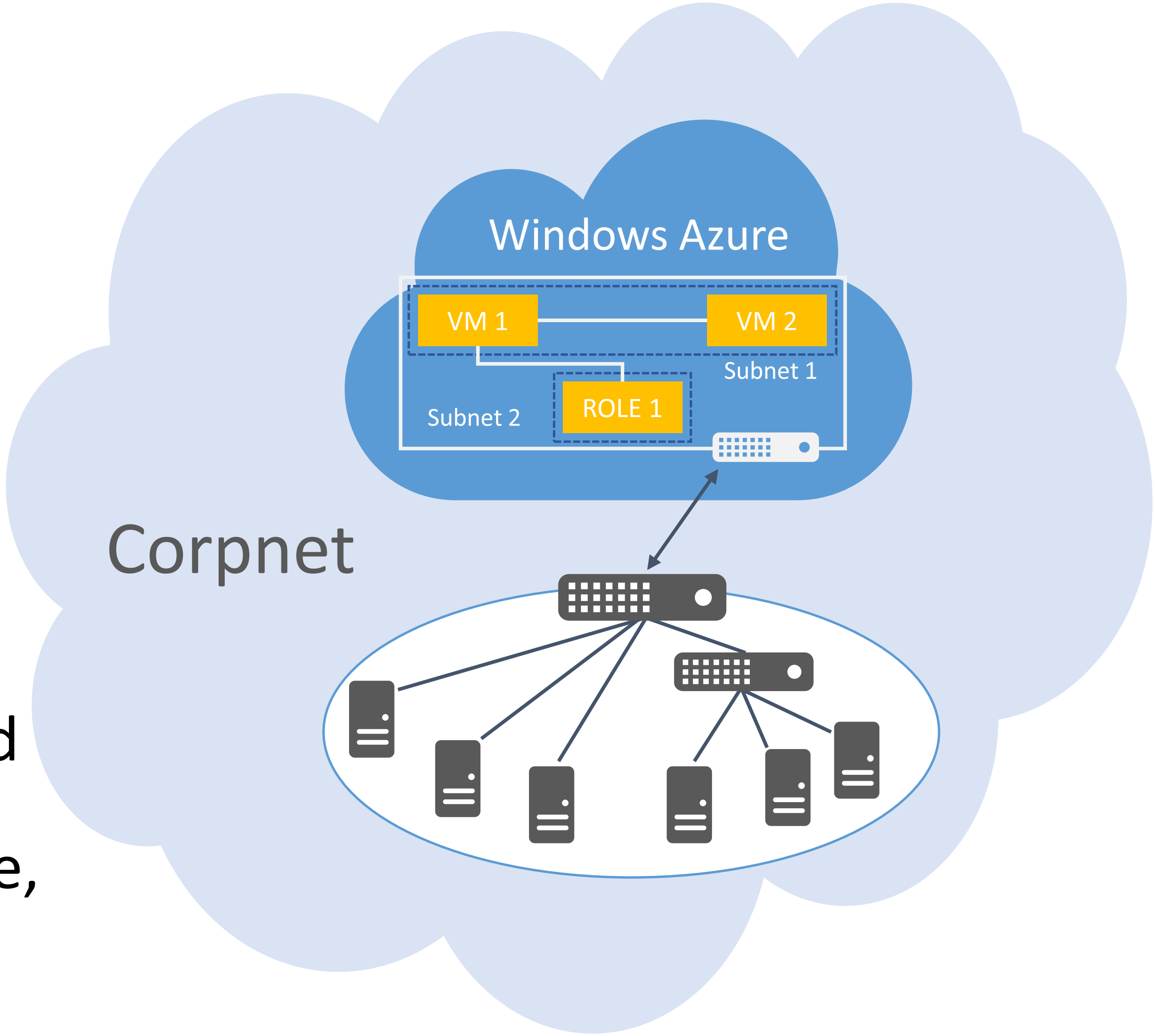


BYO Routing Policy to the Cloud

Policy lookup: 10.2/16 routes to GW on host with PA 10.1.1.7

VNet is a Hit!

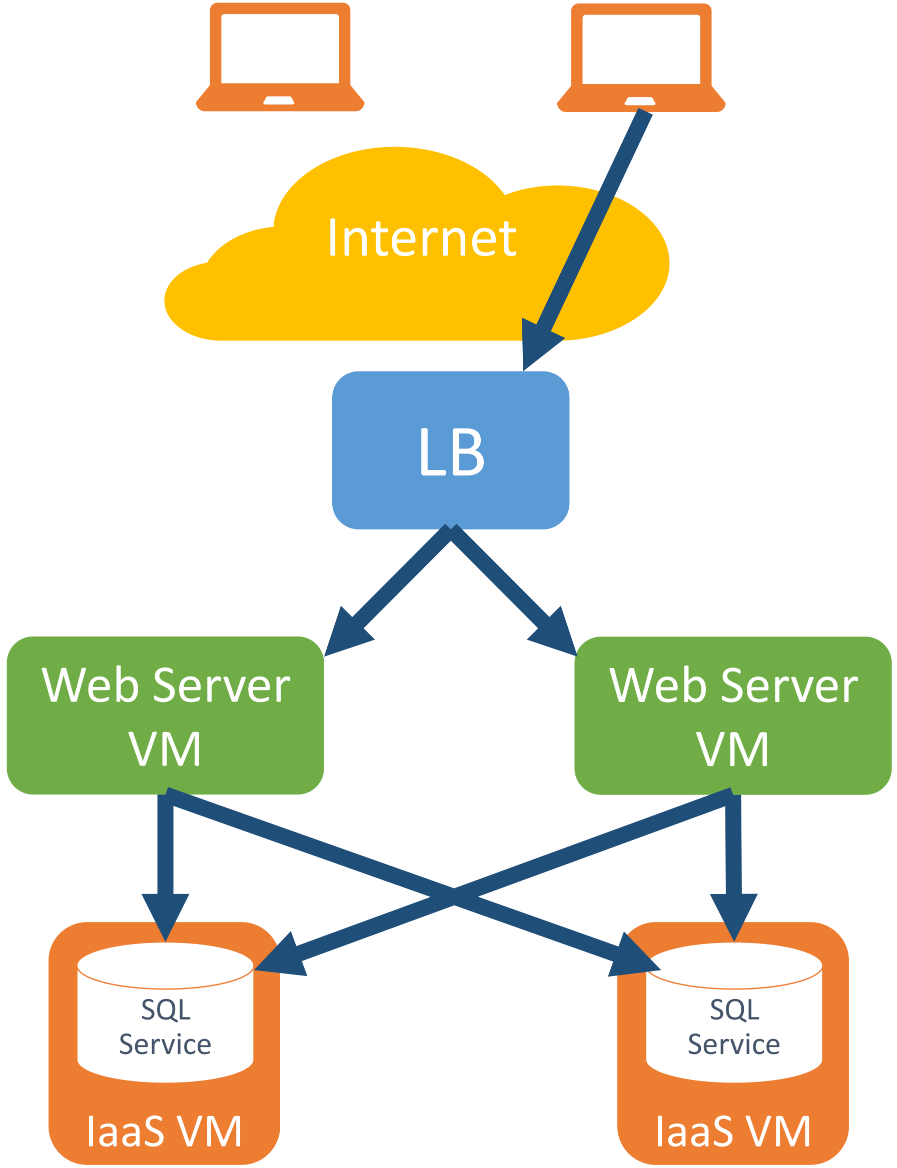
- >25% of all IaaS deployments create VNets
- Tens of thousands of VNets created since preview last summer, running gateways back to on-prem
- We have single VNets running thousands of VMs
- VNet is the central piece of our Hybrid Cloud strategy – we let enterprises migrate to the cloud at their own pace, in a familiar environment



More Cloud Networking Challenges

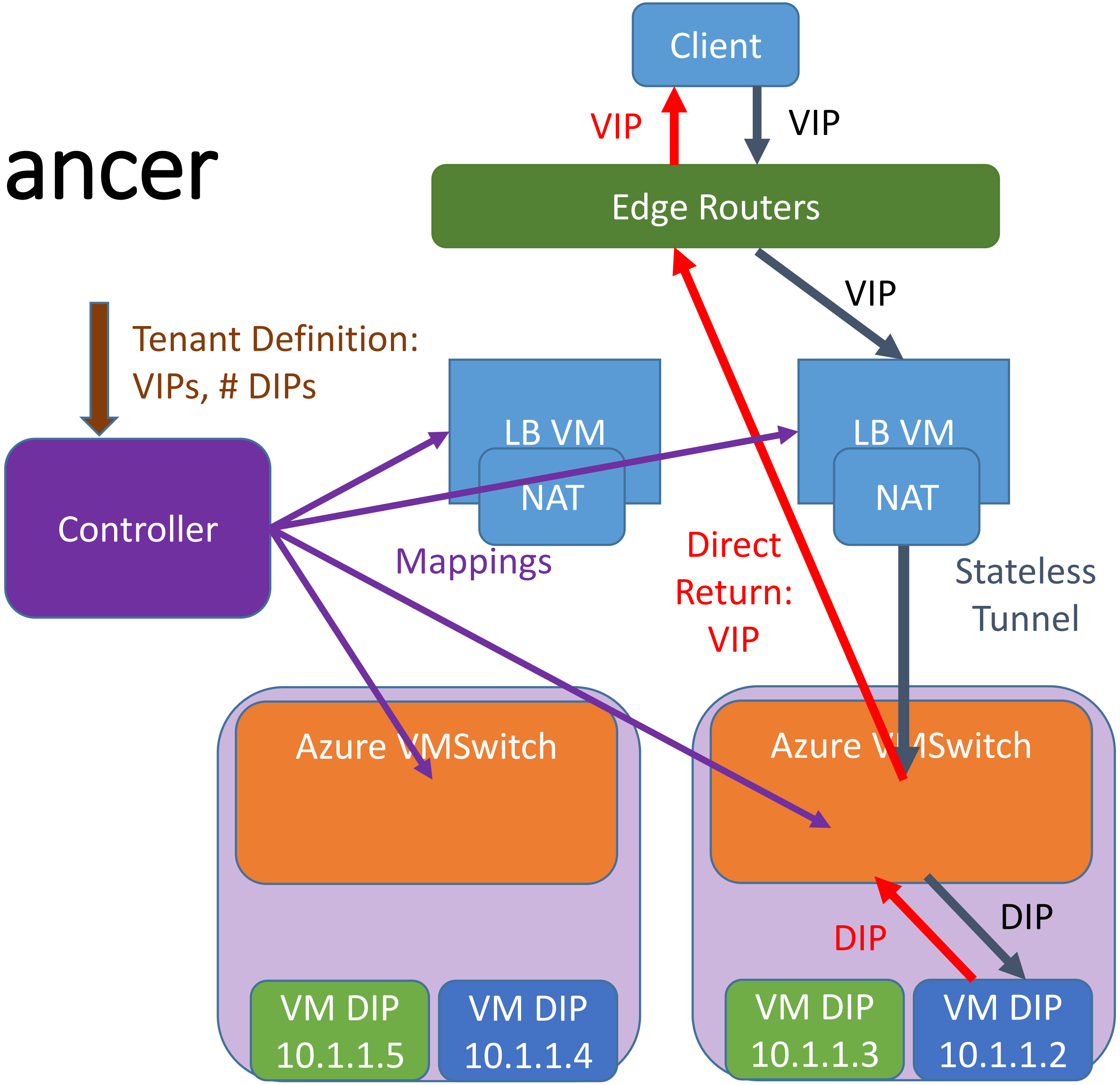
Cloud Services Run Behind Load Balancers

- All infrastructure runs behind a load balancer (LB) to enable high availability and application scale
- How do we make application load balancing scale to the cloud?
- Challenges:
  - How do you load balance the load balancers?
  - Hardware LBs are complex, and cannot support the rapid creation/deletion of LB endpoints required in the cloud
  - Support 100s of Gbps per data center
  - Need a simple provisioning model



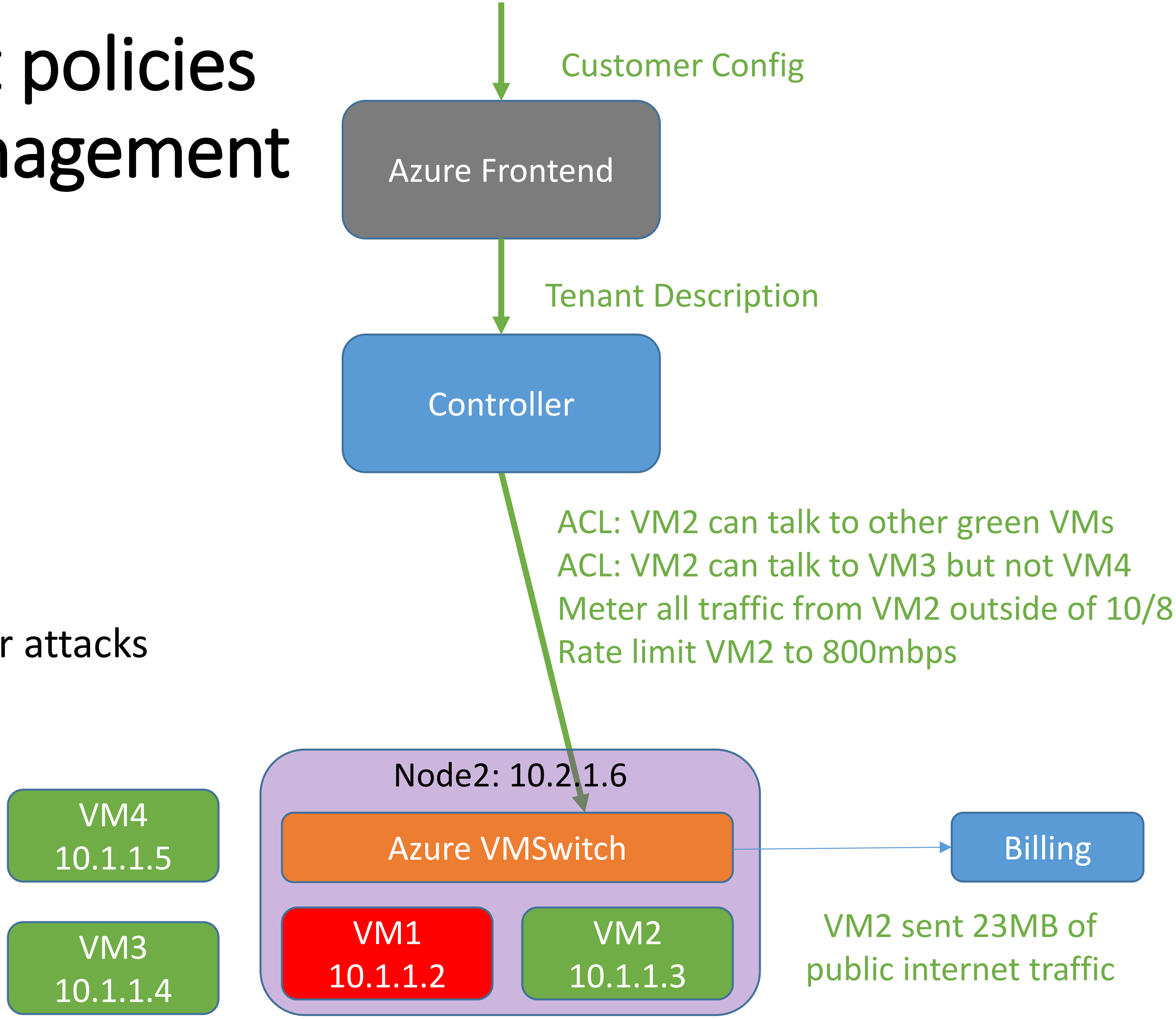
SDN Approach: All-Software Load Balancer

- Goal of an LB: Map a Virtual IP (VIP) to a set of Dedicated IP addresses (DIPs) of a cloud service
- Two steps: Load balance (select a DIP) and NAT (translate VIP→DIP and ports)
- Pushing the NAT to the VMSwitch removes state from LB and enables direct return
- Single SDN controller abstracts out LB/VMSwitch interactions



SDN abstracts all net policies ⇒ Simple policy management

- 5-tuple ACLs
  - Infrastructure Protection
  - User-defined Protection
- Billing
  - Metering traffic to internet
- Rate limiting
- Security Guards
  - Spoof, ARP, DHCP, and other attacks
- Per-tenant DNS
- VLANs
- Physical switches
- More in development…



Customer Config

Azure Frontend

Tenant Description

Controller

ACL: VM2 can talk to other green VMs
ACL: VM2 can talk to VM3 but not VM4
Meter all traffic from VM2 outside of 10/8
Rate limit VM2 to 800mbps

Node2: 10.2.1.6

Azure VMSwitch

VM1
10.1.1.2

VM2
10.1.1.3

VM4
10.1.1.5

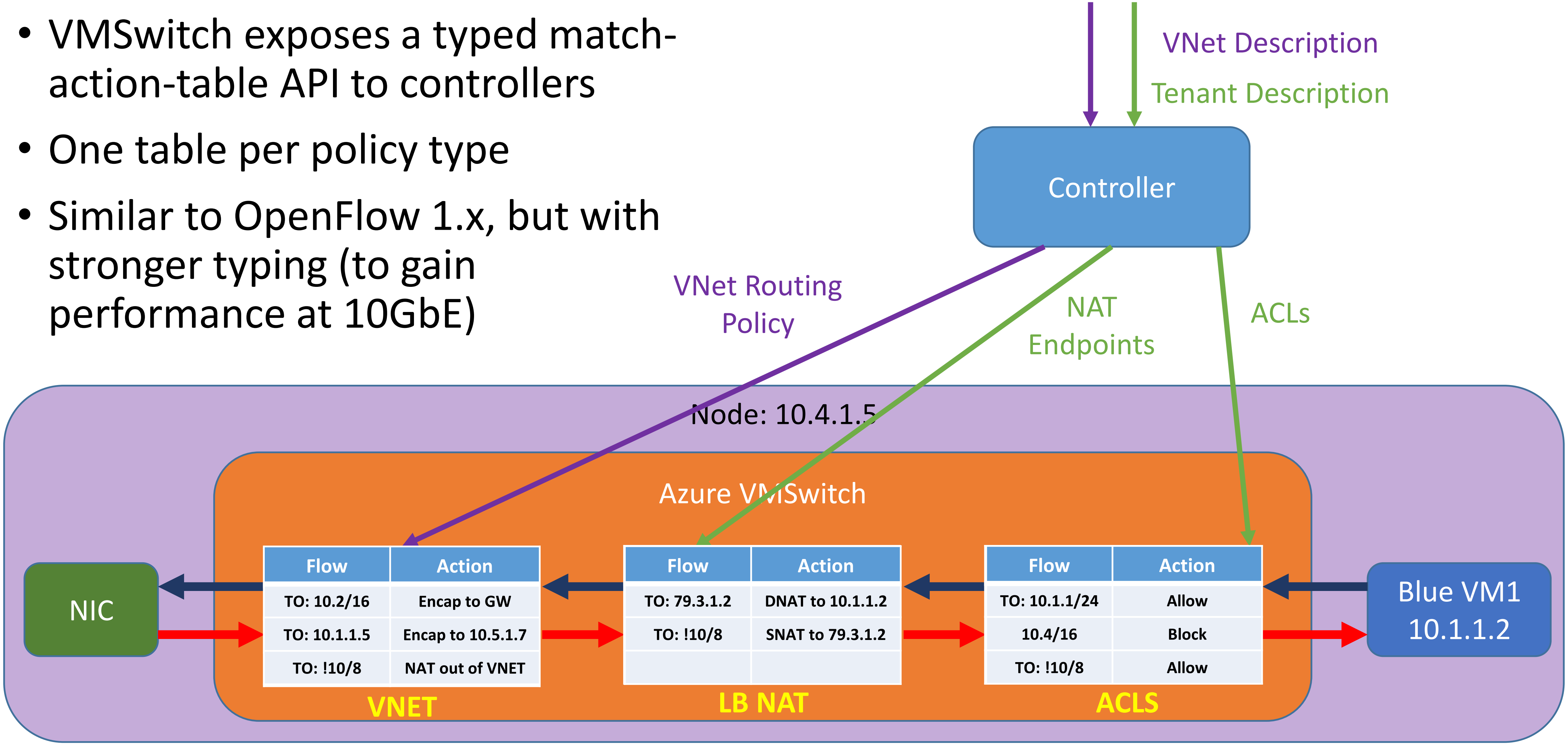
VM3
10.1.1.4

Billing

VM2 sent 23MB of
public internet traffic

VMSwitch: The intersection of all policy

- VMSwitch exposes a typed match-action-table API to controllers
- One table per policy type
- Similar to OpenFlow 1.x, but with stronger typing (to gain performance at 10GbE)



VNET

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

LB NAT

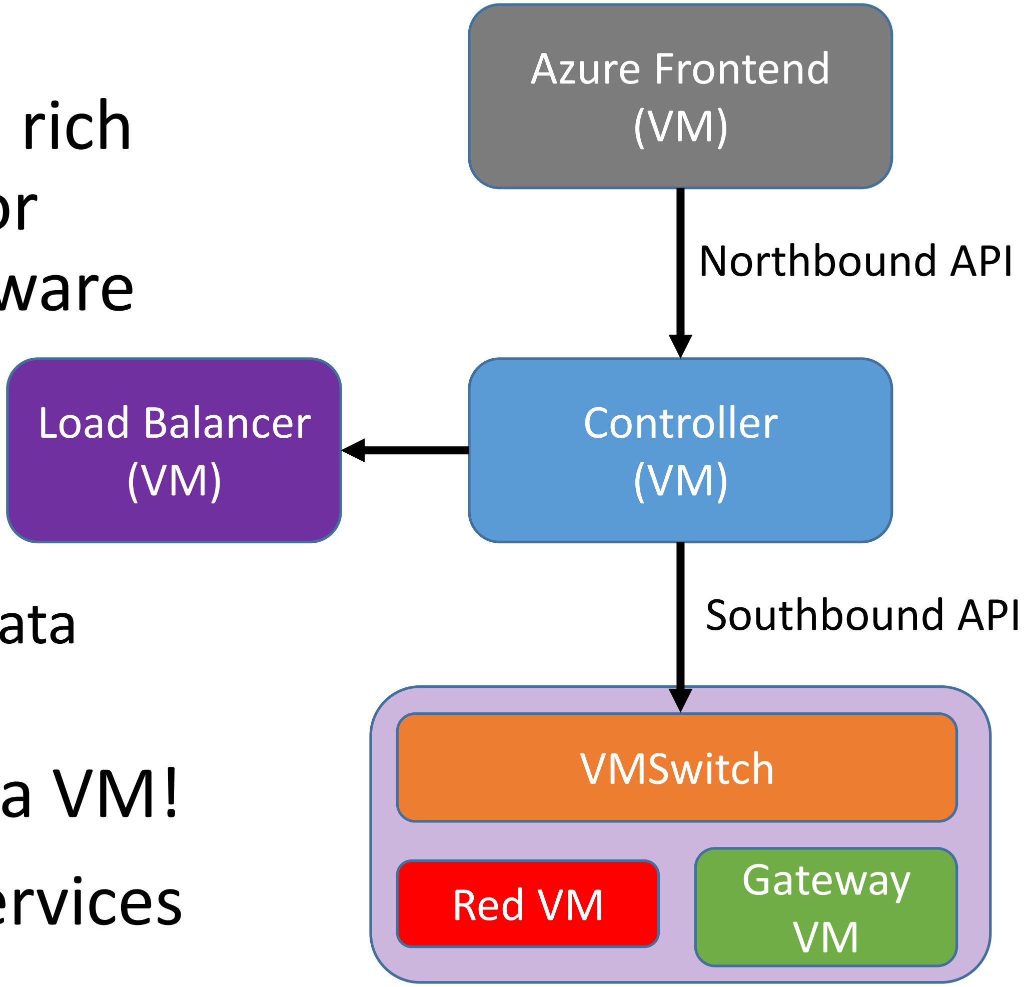
Flow	Action
TO: 79.3.1.2	DNAT to 10.1.1.2
TO: !10/8	SNAT to 79.3.1.2

ACLS

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

End Result: Agility and Scale

- Windows Azure supports virtual networks, rich load balancing, tenant ACLs, and more – for hundreds of thousands of servers, via software
  - No Hardware per tenant ACLs
  - No Hardware NAT
  - No Hardware VPN / overlay
  - No Vendor-specific control, management or data plane
- All policy is in software – and everything's a VM!
- Network services deployed like all other services



We bet our infrastructure on SDN, and it paid off

Final thoughts: Challenges Overcome

- It's not scalable to synchronously push the entire network state to every VMSwitch
  - We have to enable eventual consistency on the network with event-driven control plane updates
  - A wire protocol back to the controller doesn't scale – we need smart agents
- The VMSwitch is policy rich, but needs to support 10GbE/40GbE
  - Extensive flow hashing is required, as well as strict table typing
  - Managing latency and latency jitter is getting increasingly difficult
- SDN scales better, but it still has finite scale; need to federate controllers
  - It's a challenge to federate the controller sets and still achieve consistent policy
  - Have to federate the Northbound API