CS244

Advanced Topics in Networking

Lecture 9: SDN (2)

Network Virtualization

Nick McKeown

"Network Virtualization in Multi-tenant Datacenters,"
[Teemu Koponen et al, 2014]

Spring 2020

Context

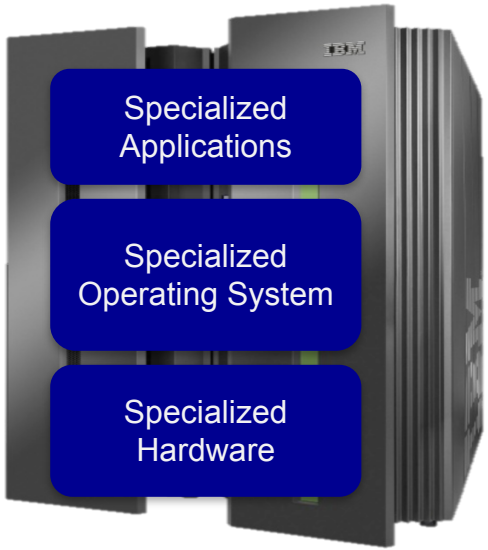
Teemu Koponen

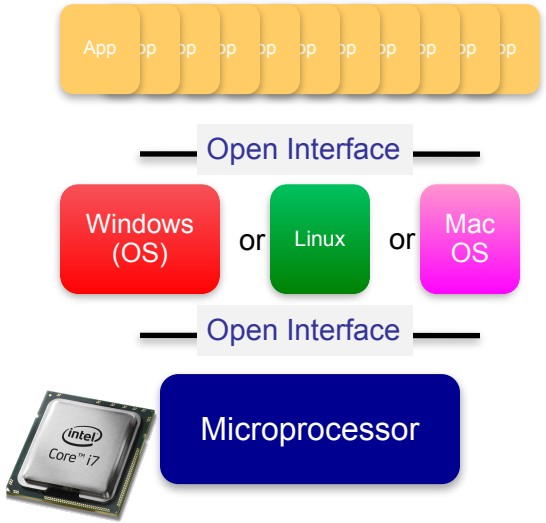
- Early employee Nicira
- Sigcomm Rising Star Award, 2012
- More recently, co-founder at Styra

Teemu

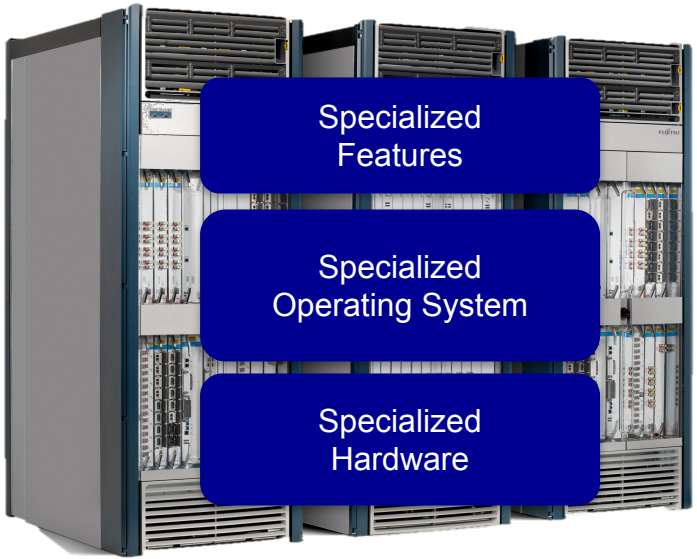
SDN: In the context of bigger networking industry changes

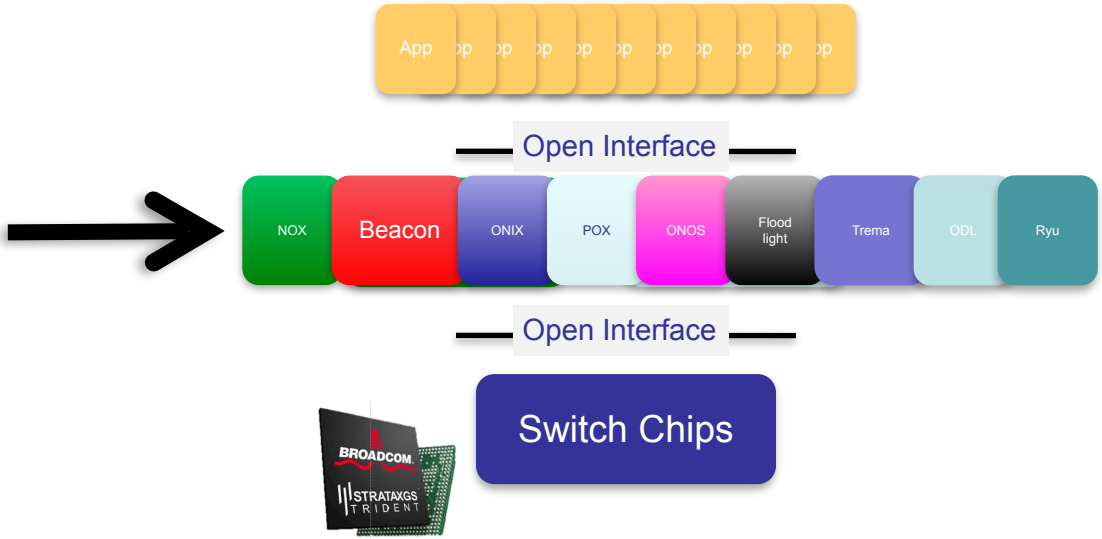
Computer Industry





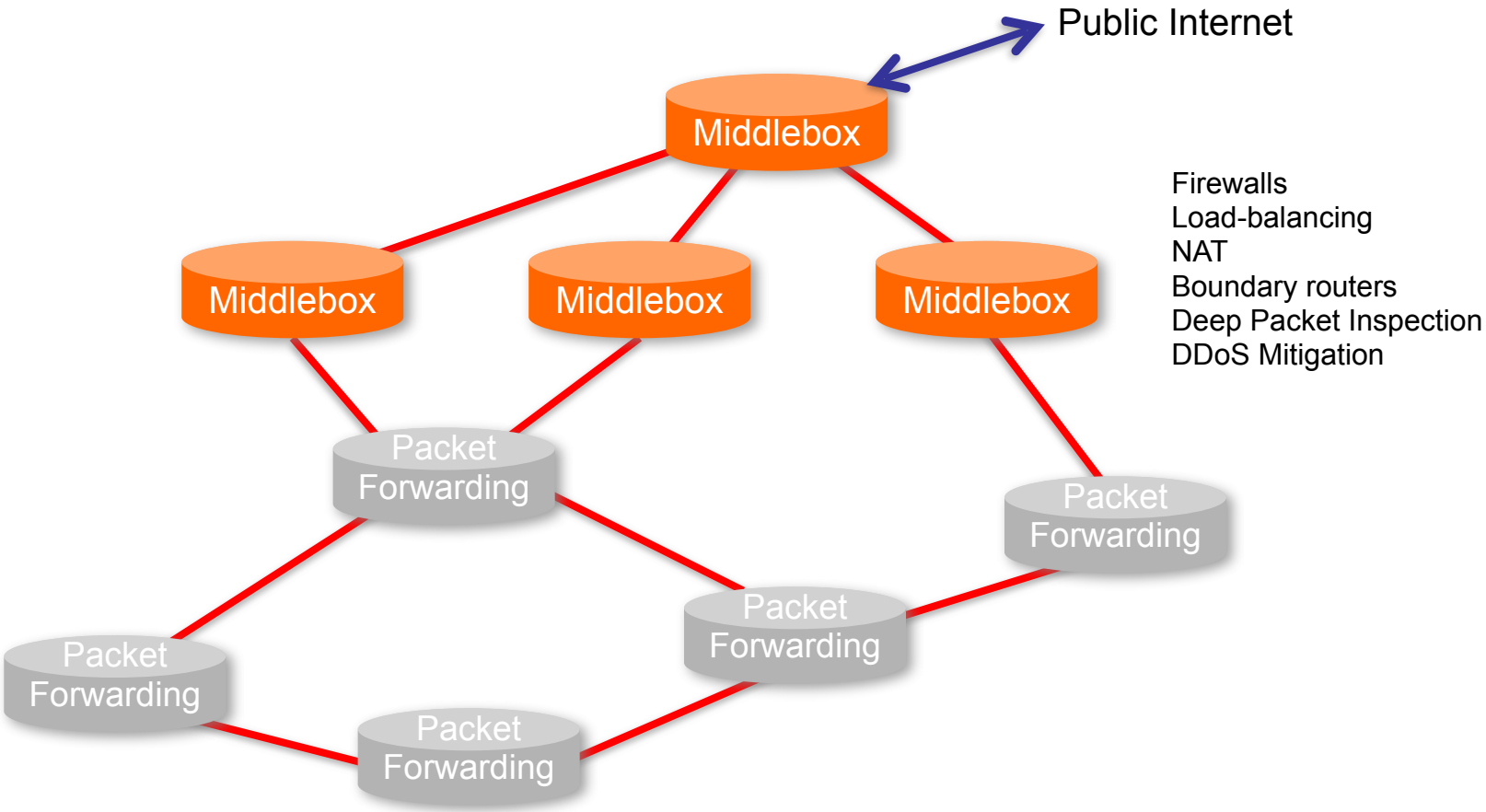
Networking Industry



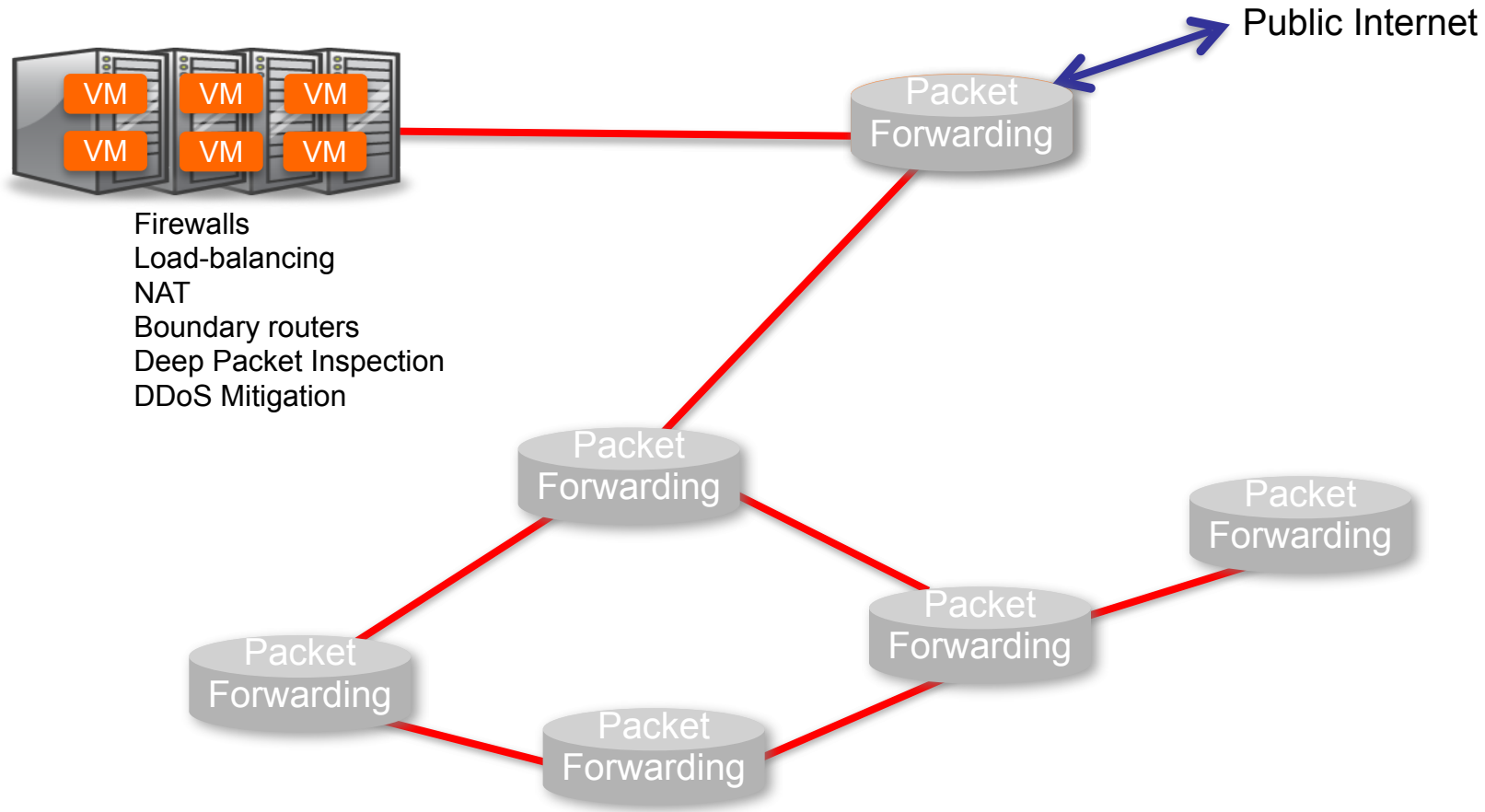


"Software is eating the world (of networking)"

Network Function Virtualization (NFV)



Network Function Virtualization (NFV)



With hindsight, Disaggregation, SDN and NFV were probably inevitable

Part of a bigger trend towards the owners and operators of networks taking control of how they manage their networks

Inevitable because…

1. Rise of Linux.
2. Rise of baremetal servers and data centers.
3. SDN: Rise of merchant switching silicon.
4. NFV: Rise of computer virtualization.

Today

Most networking equipment is disaggregating

- Intra- and inter-datacenter networks
- ISP routers and switches
- WiFi APs
- Cellular basestations (4G, 5G…)
- Optical and Metro Transport
- Residential broadband access
- Enterprise network equipment: switch, router, firewall

Network Virtualization

“Modularity based on abstraction is the way things are done!”

Barbara Liskov (MIT)
Turing Award Lecture 2009

Abstractions in computer systems

Virtual memory: Abstract illusion of infinite, private physical memory

File system: Uniform illusion of read/write data store.

Virtual Machine: User application cannot tell if it is running on a physical or virtual machine.

…

What is “network virtualization”?

In this context: The abstraction (or illusion) of a physical network in which the user, application (and possibly the administrator too) cannot tell if the network is physical or virtual.

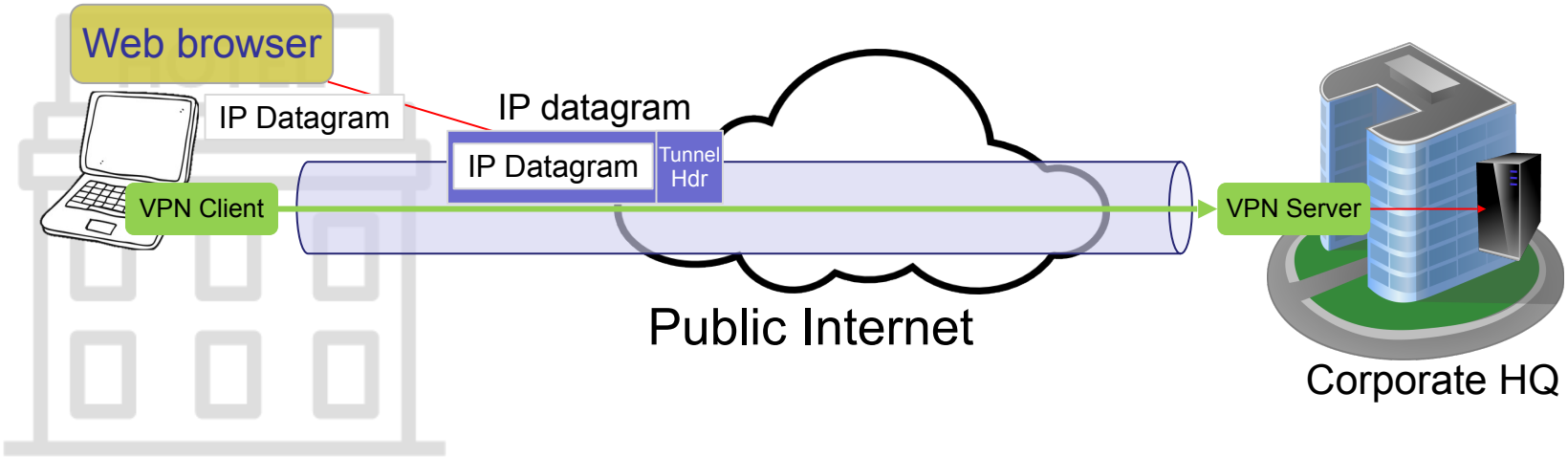
Q: If true, what would be the benefits?

Will Robert Brand:

…does this kind of virtualization have any advantages in aggregate? That is, under NVP, are there positive or negative externalities to running diverse logical topologies in the same datacenter?

Early attempts at network virtualization

Example: VPN

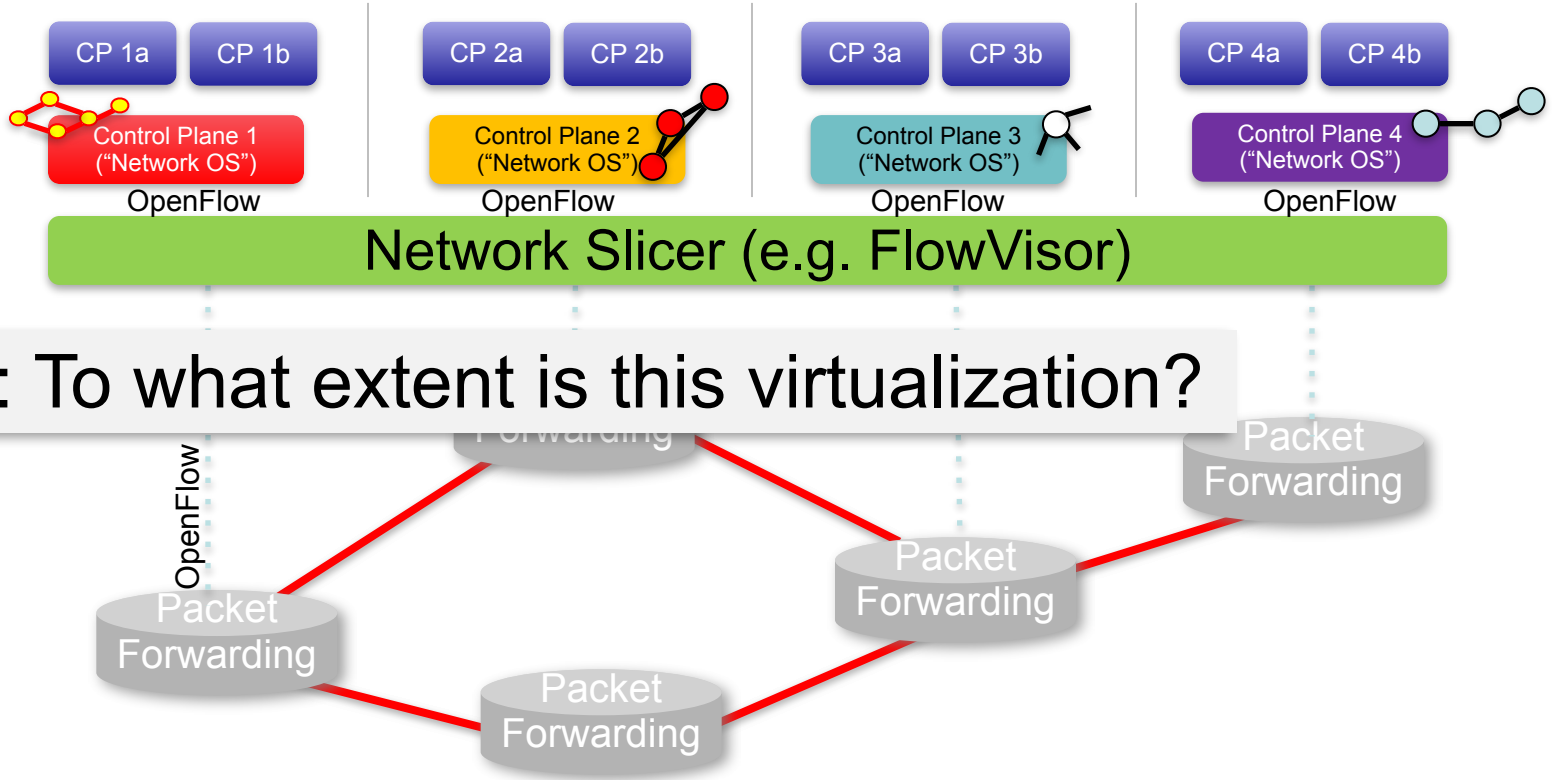


Q: To what extent is this virtualization?

Early attempts at network virtualization

Example: Slicing

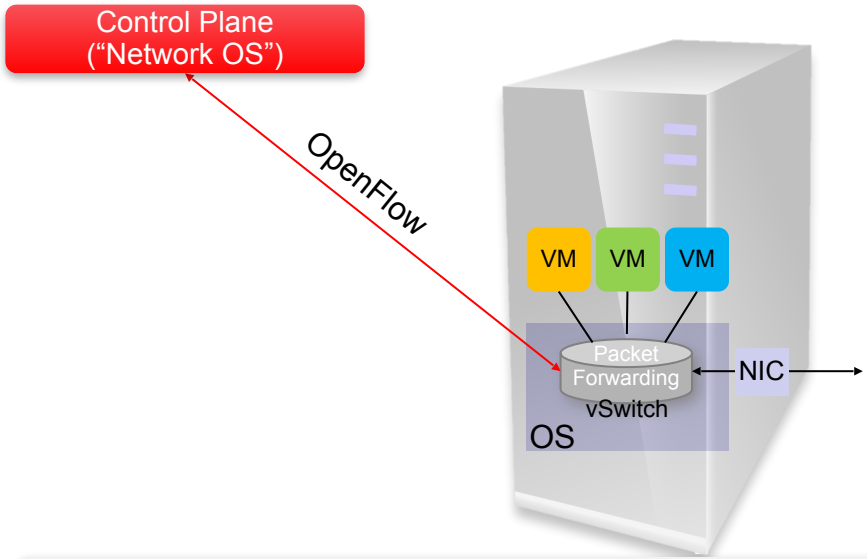
Each controller can read and/or write flow rules for its assigned portion of “header space” and topology



Trends at the time of writing

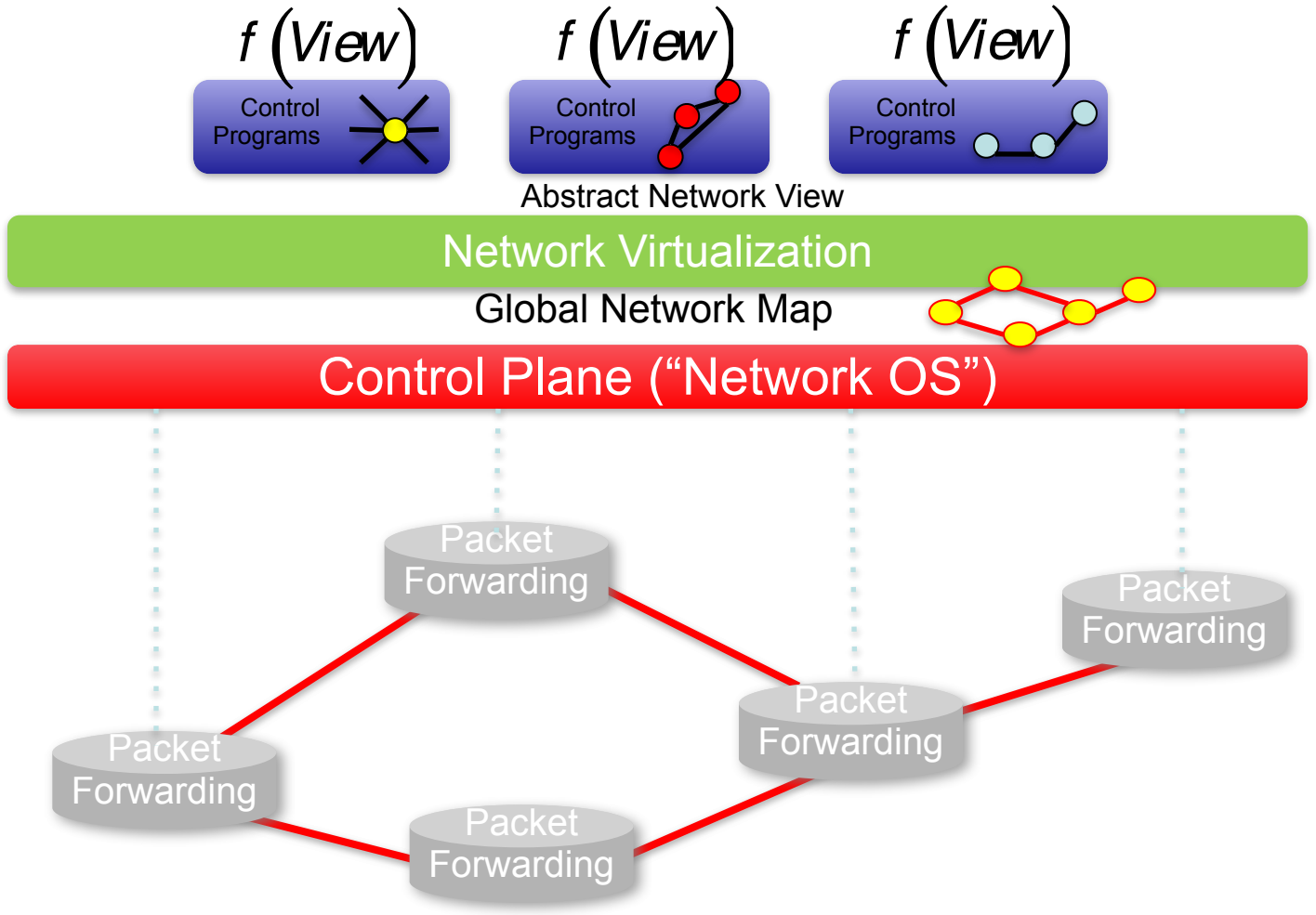
1. Data centers and clouds ☞ efficiency matters
2. VMs ☞ a vSwitch inside every server
3. SDN ☞ abstraction for control

Virtual vSwitch in every server

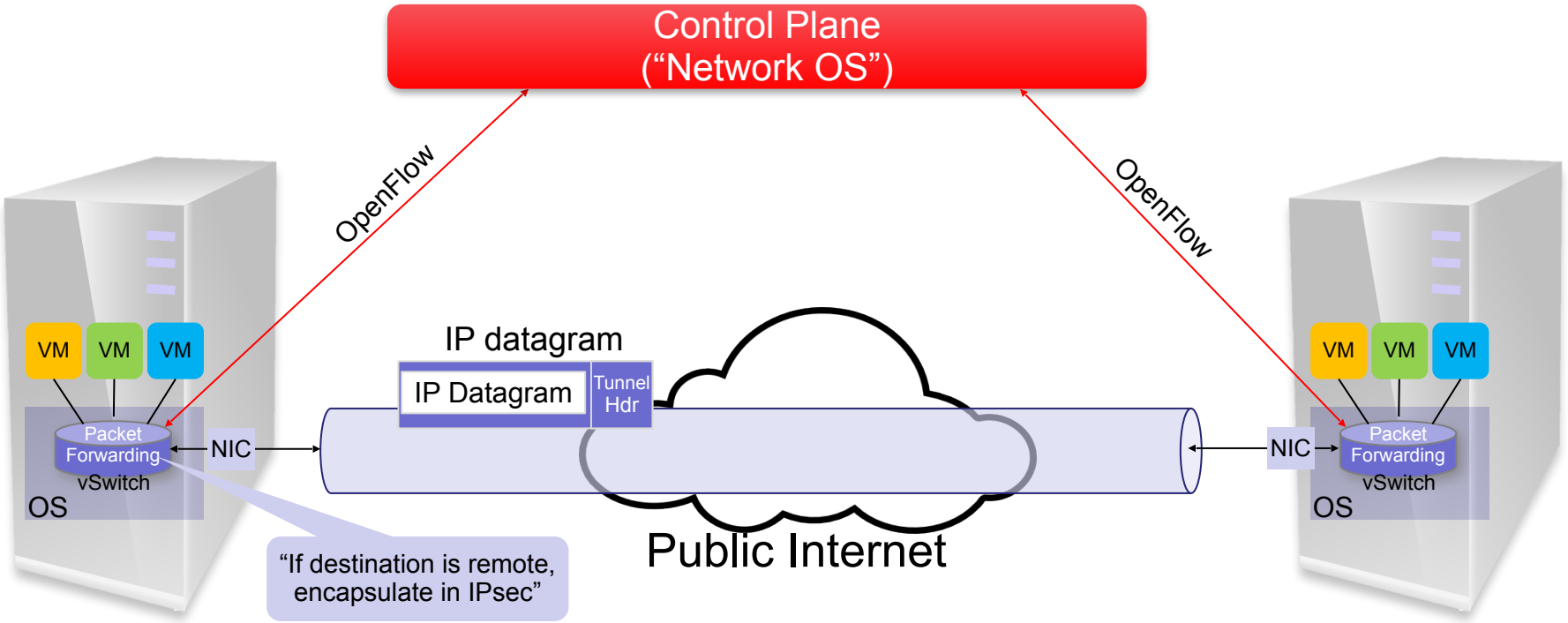


Q: How might the vSwitch help?

SDN and Network Virtualization

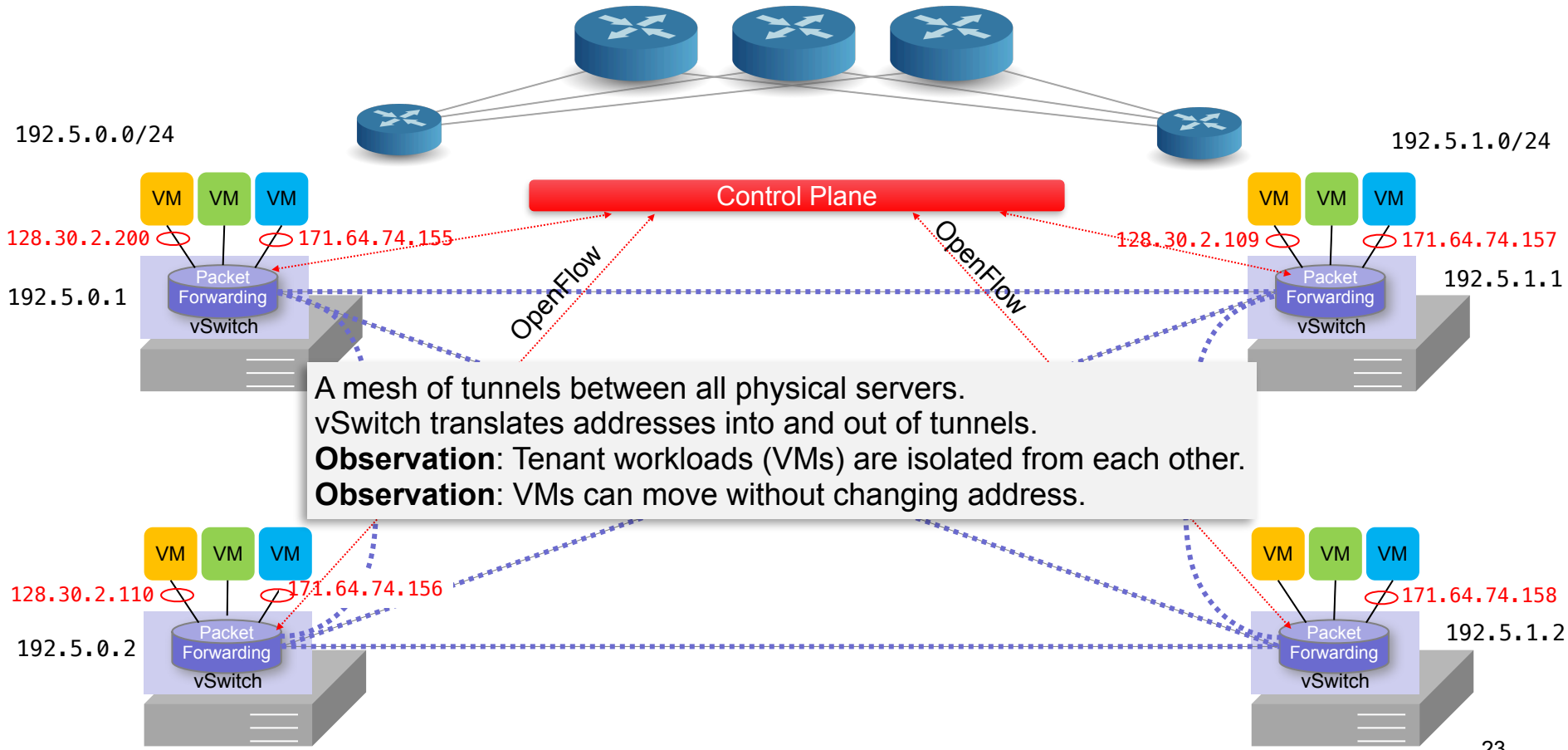


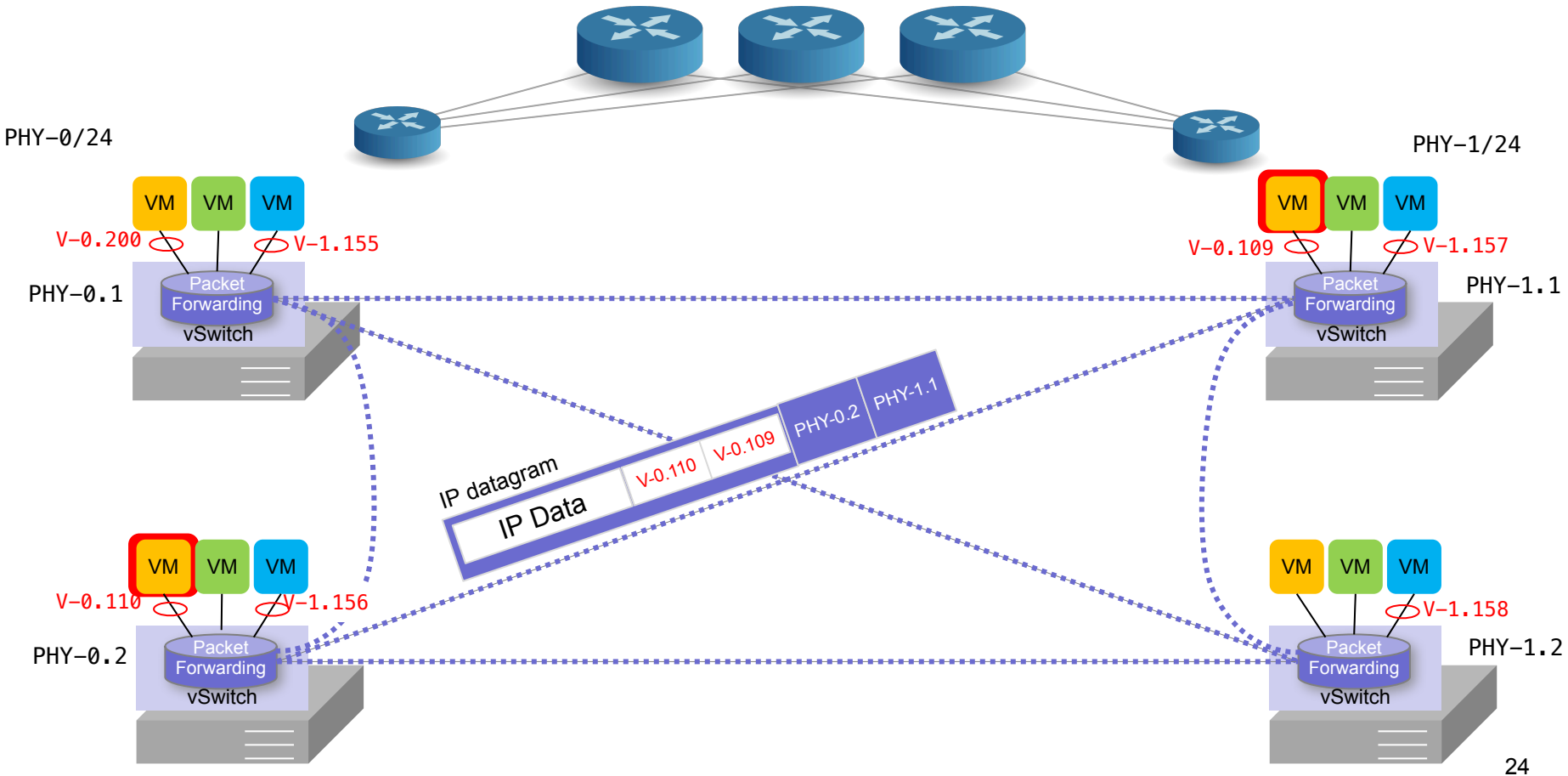
Another way to create a VPN

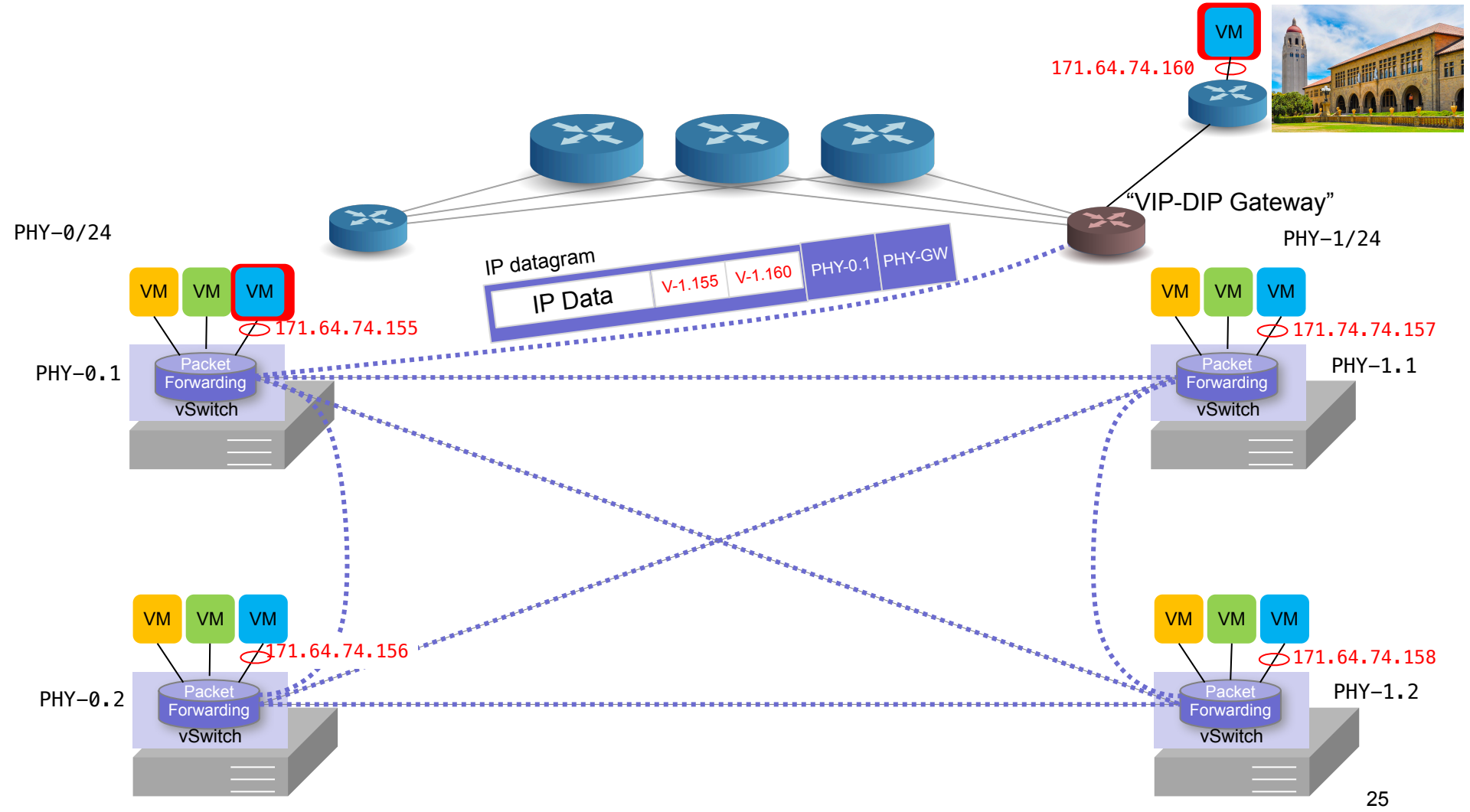


Observation 1: Control Plane tells vSwitch how to process packets into/out of tunnel

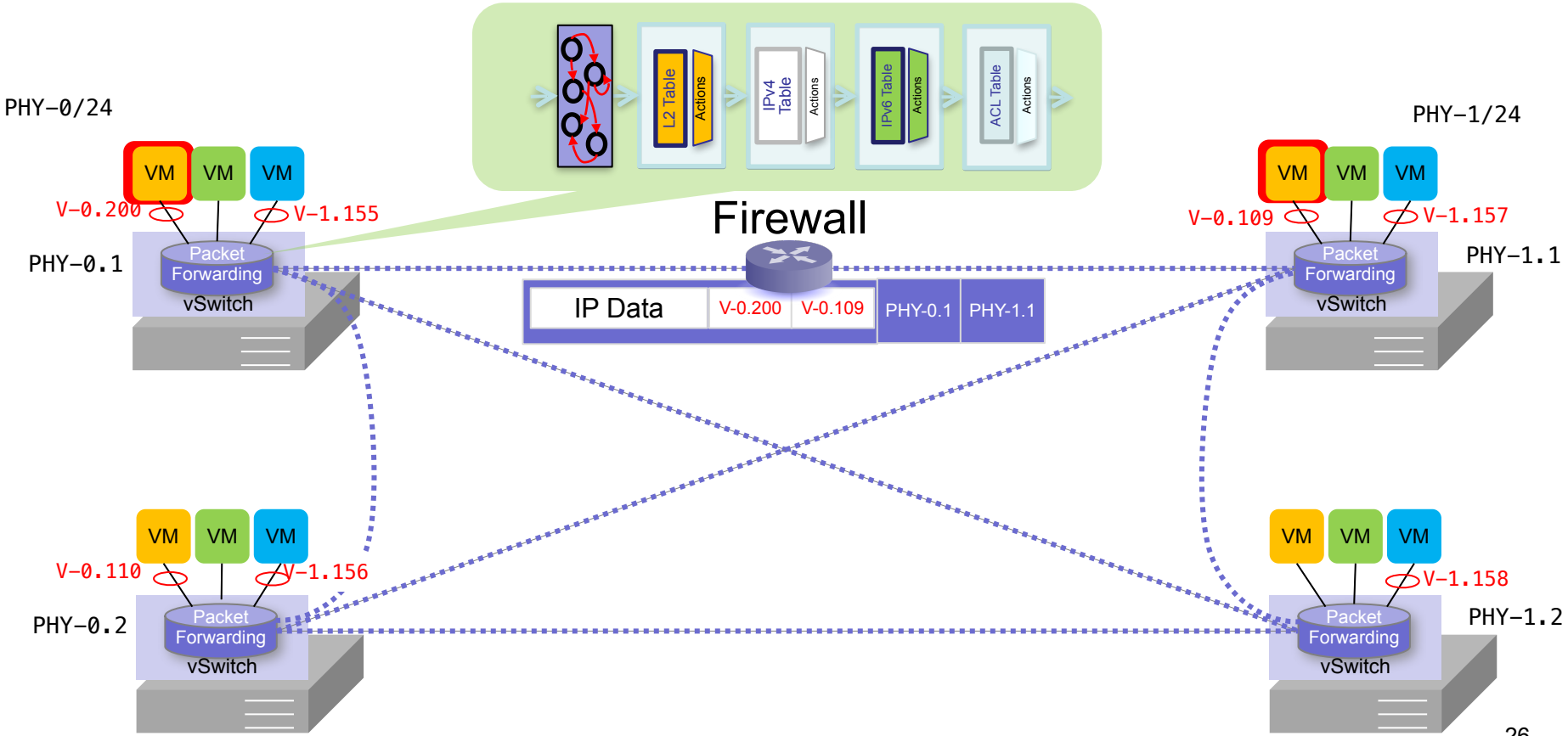
In a virtualized cloud service provider



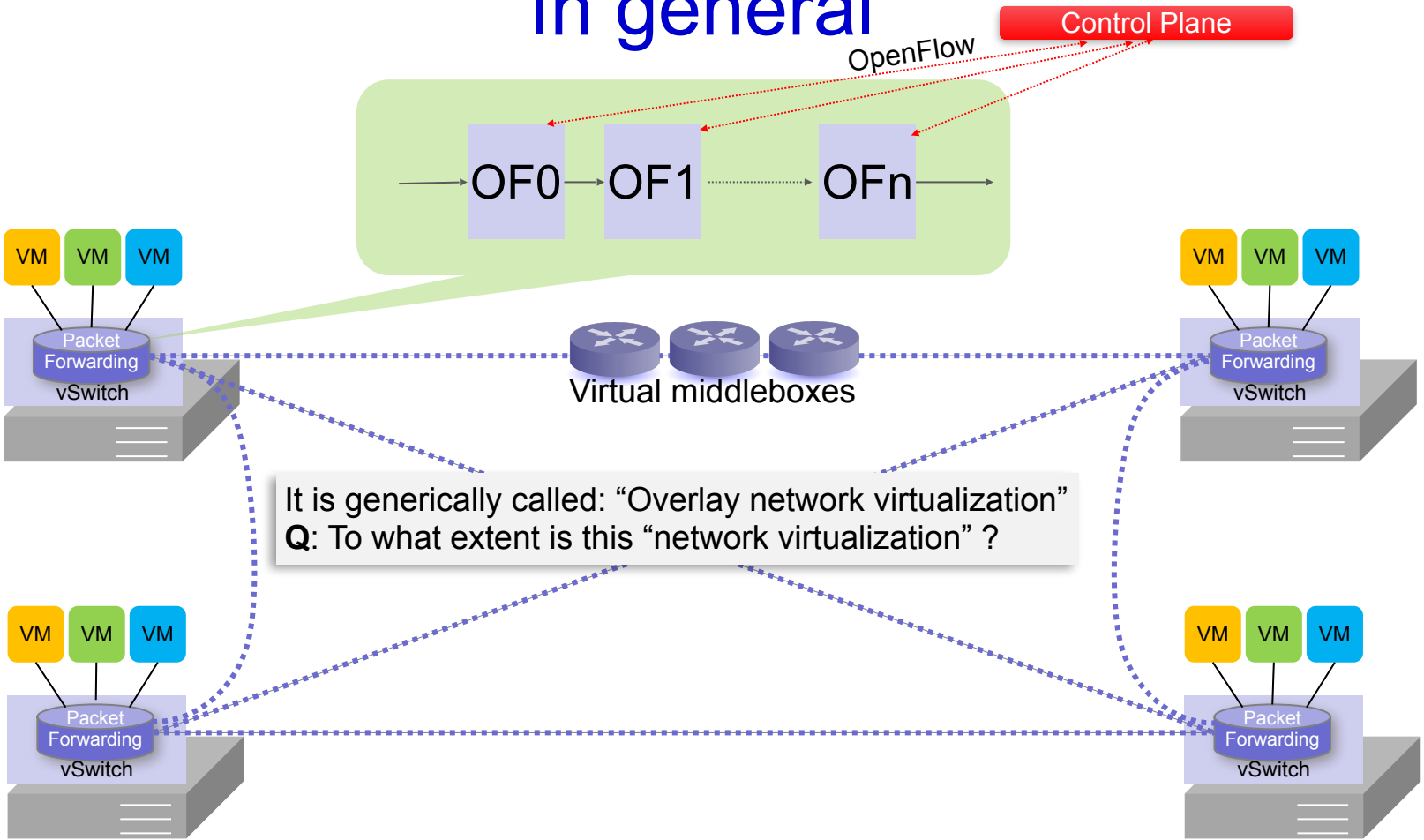

PHY-0/24
PHY-1/24
VM
VM
VM
V-0.200
V-1.155
PHY-0.1
Packet Forwarding
vSwitch
V-0.109
V-1.157
PHY-1.1
IP datagram
IP Data
V-0.110
V-0.109
PHY-0.2
PHY-1.1
V-0.110
V-1.156
PHY-0.2
V-1.158
PHY-1.2


VM
171.64.74.160
"VIP-DIP Gateway"
PHY-0/24
PHY-1/24
IP datagram
IP Data
V-1.155
V-1.160
PHY-0.1
PHY-GW
VM
VM
VM
171.64.74.155
PHY-0.1
Packet Forwarding
vSwitch
VM
VM
VM
171.74.74.157
PHY-1.1
Packet Forwarding
vSwitch
VM
VM
VM
171.64.74.156
PHY-0.2
Packet Forwarding
vSwitch
VM
VM
VM
171.64.74.158
PHY-1.2
Packet Forwarding
vSwitch

Adding a distributed, virtual firewall



In general

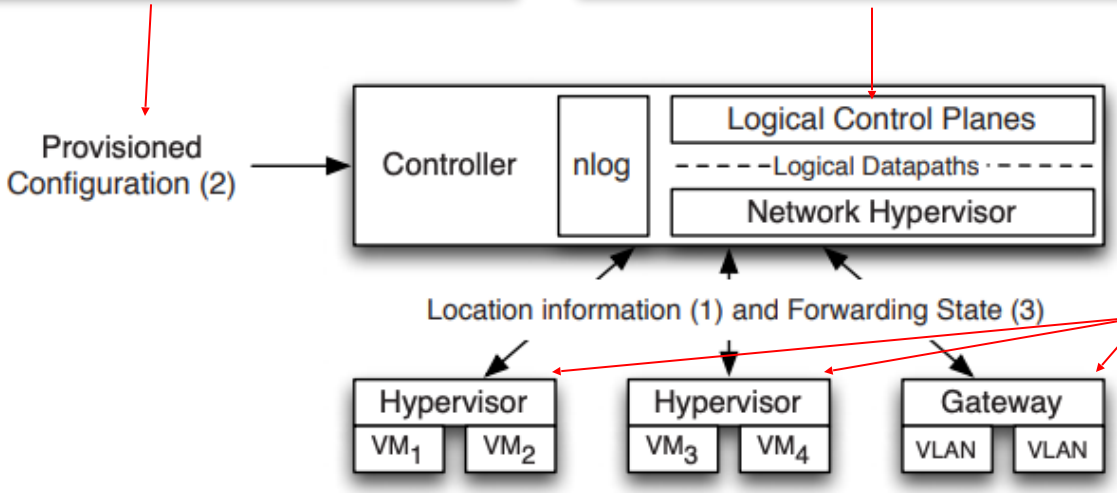


It is generically called: “Overlay network virtualization”
Q: To what extent is this “network virtualization” ?

NVP is proactive: Pushes rules and state top-down

2: Datacenter owner configures the networks: topologies and protocols.

3: Calculates forwarding pipeline model for each vSwitch and the state for each table. Pushes via OpenFlow and OVS control protocol.



1: Provide location, state and topology. Even as VMs move. OVS control protocol.

Figure 5: Inputs and outputs to the forwarding state computation process which uses nlog, as discussed in §4.3.

Top-down proactive control

Goals:

- **Scale**: Controller does not process packets
- **Isolation**: To continue if VMs, vSwitches fail ☞ controller is a distributed, resilient cluster

Immense computational challenge!

- NVP is built on ONIX, a distributed SDN controller (used by Google)
- Each NVP/ONIX controller manages some slices (shards); and is responsible for others, if a controller fails.
- NVP uses Apache Zookeeper for
  - Leader election: to coordinate global resources and load-balancing) and
  - Label allocation: Logical egress port must be globally unique

Scaling Challenges

Margalit Ruth Glasgow:

Technical question: The authors mention in the "Lessons Learned" that using OpenFlow requires O(n^2) operations to tailor flows for each hypervisor, vs the standard complexity of O(n) for the logical controller. Where do those numbers come from? Does the O(n) come from each VM having O(1) connections at once, and the O(n^2) from each hypervisor needing to always have a connection with each other hypervisor?

A: Yes.

Q: What did the authors plan for the next version of NVP?

Neil Perry

This came out in 2014 (NSDI '14). Have any "simpler" systems that achieve the same goals as this been put forward since? NVP seems very complicated and hard to implement (lots of room for mistakes).

Ryan Smith:

How can you implement a similar concept without using a centralized OpenFlow-like model to avoid the scaling penalty?

End.