cloudstack

Xen*, SDN and Apache Cloudstack

Sebastien Goasguen,
Apache CloudStack Citrix EMEA

August 28th 2012

Xen Summit

cloudstack

Outline

- A bit about CloudStack
- A bit about SDN
- A bit about OpenVswitch
- Some bits about “SDN” in CloudStack

- Slides are on slideshare for download:

http://www.slideshare.net/sebastiengoasguen/cloudstack-and-sdn

cloudstack

Apache CloudStack

- IaaS solution to build a private/public cloud
- Hypervisor agnostic:
  - Xen, XS, XCP
  - KVM
  - VMware
- Object store support
  - Swift
  - Upcoming support from Caringo
- EC2/S3 compatibility

- Java application
- Ant build but moving to maven (via new contributor)

- First Apache release 4.0 coming Sept 26th

cloudstack

Participating in CloudStack

- Apache incubator project
- http://www.cloudstack.org
- #cloudstack on irc.freenode.net
- @CloudStack on Twitter
- http://cloudstack.org/discuss/mailing-lists.html

Welcoming contributions and feedback, Join the fun !

cloudstack

A Very Flexible IaaS Platform

Compute

Hypervisor

- XenServer
- VMware
- Oracle VM
- KVM
- Bare metal

Storage

Block & Object

- Local Disk
- iSCSI
- Fiber Channel
- NFS
- Swift

Primary Storage

Secondary Storage

Network

Network & Network Services

- Network Type
- Isolation
- Firewall
- Load balancer
- VPN

http://www.slideshare.net/cloudstack/cloudstack-architecture

cloudstack

NaaS ?

- “Cloud Servers”
  - On-demand, Elastic, Measured server provisioning
- Cloud Storage
  - Scalable/fault tolerant storage with object stores
- Cloud Networks
  - How to do on-demand, elastic, measured networking provisioning ?
  - **How to program the network ?**

cloudstack

A very extensive API

Network

- createNetworkOffering
- updateNetworkOffering
- deleteNetworkOffering
- listNetworkOfferings
- createNetwork
- deleteNetwork (A)
- listNetworks
- restartNetwork (A)
- updateNetwork (A)
- createPhysicalNetwork (A)
- deletePhysicalNetwork (A)
- listPhysicalNetworks
- updatePhysicalNetwork (A)
- listSupportedNetworkServices
- addNetworkServiceProvider (A)
- deleteNetworkServiceProvider (A)
- listNetworkServiceProviders
- updateNetworkServiceProvider (A)
- createStorageNetworkIpRange (A)
- deleteStorageNetworkIpRange (A)
- listStorageNetworkIpRange
- updateStorageNetworkIpRange (A)
- addNetworkDevice
- listNetworkDevice
- deleteNetworkDevice
- listF5LoadBalancerNetworks
- listSrxFirewallNetworks
- listNetscalerLoadBalancerNetworks

Load Balancer

- createLoadBalancerRule (A)
- deleteLoadBalancerRule (A)
- removeFromLoadBalancerRule (A)
- assignToLoadBalancerRule (A)
- createLBStickinessPolicy (A)
- deleteLBStickinessPolicy (A)
- listLoadBalancerRules
- listLBStickinessPolicies
- listLoadBalancerRuleInstances
- updateLoadBalancerRule (A)
- addF5LoadBalancer (A)
- configureF5LoadBalancer (A)
- deleteF5LoadBalancer (A)
- listF5LoadBalancers
- addNetscalerLoadBalancer (A)
- deleteNetscalerLoadBalancer (A)
- configureNetscalerLoadBalancer (A)
- listNetscalerLoadBalancers

- CloudStack orchestrates your network:
  - Provisioning
  - Configuration
  - Updates
- For multi-tenants isolation
- Using hardware and software devices
- At scale: O(10^4) Hyp, O(10^5) VMs…

cloudstack

Software Defined Networking

- Enable innovation, experimentation, optimization and customization of networks
- Move control of the network to software. i.e **Programmable network**
- **Virtualize the network**
- Vendor-agnostic, standard protocol for control: OpenFlow

OpenFlow

- Google achieved 95% utilization of WAN backbone by using SDN

- Leading SDN protocol
- Decouples control and data plane by giving a controller the ability to install flow rules on switches.
- Hardware or software switches can use OpenFlow
- Spec driven by ONF

OpenFlow

- OpenFlow rules can *drop*, *rewrite*, *forward* packets

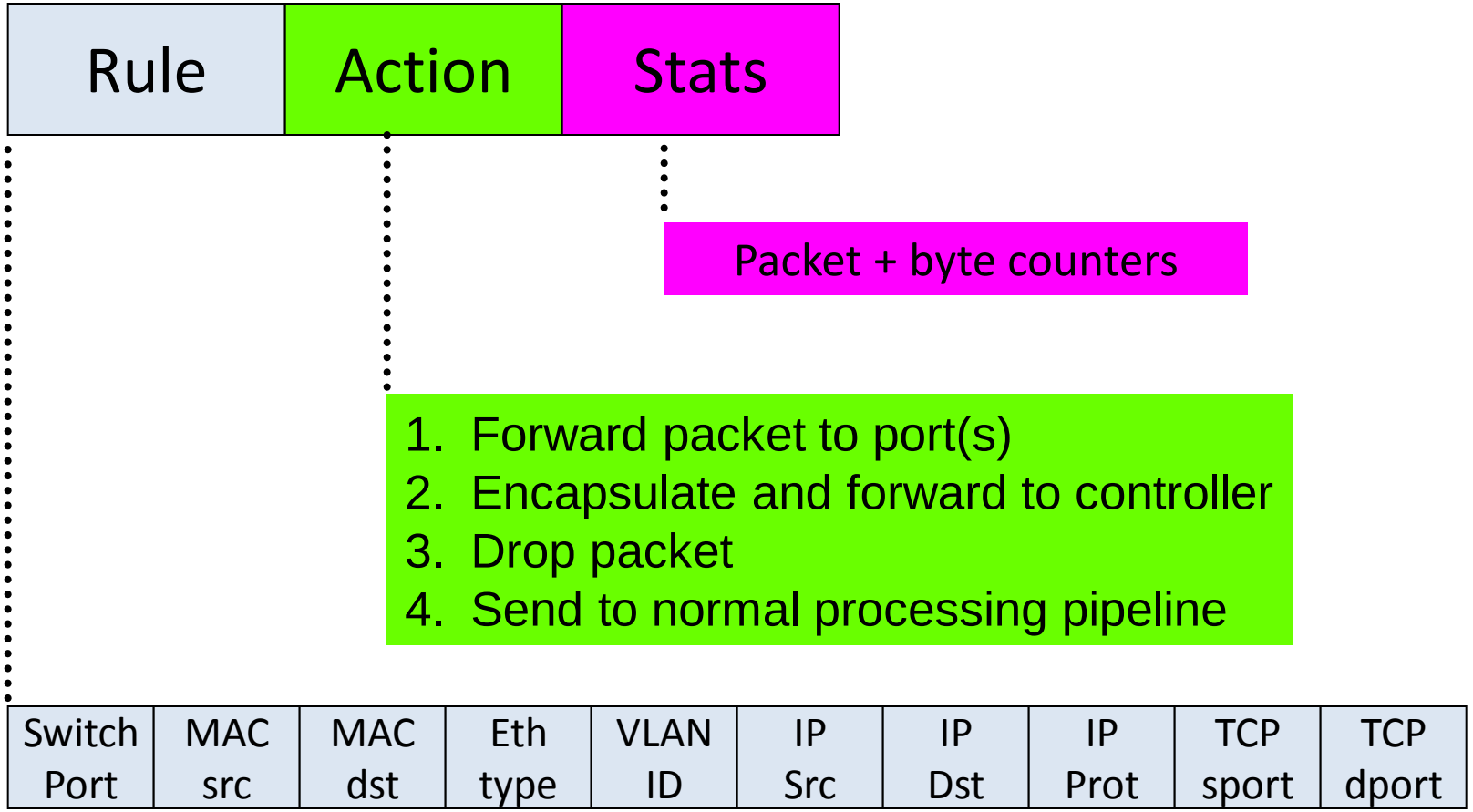
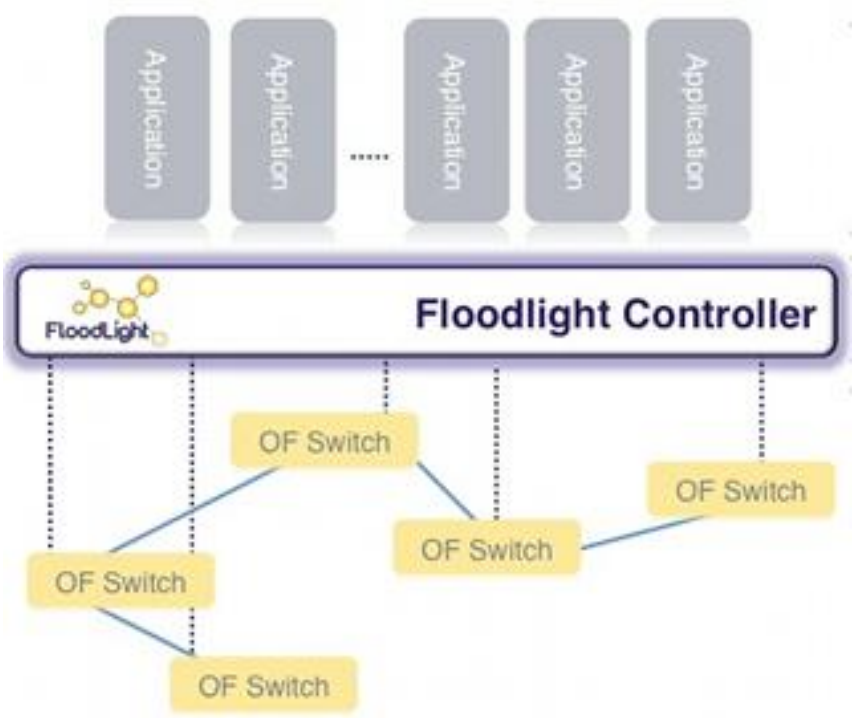


Diagram Src: http://www.openflow.org/wp/documents/

cloudstack

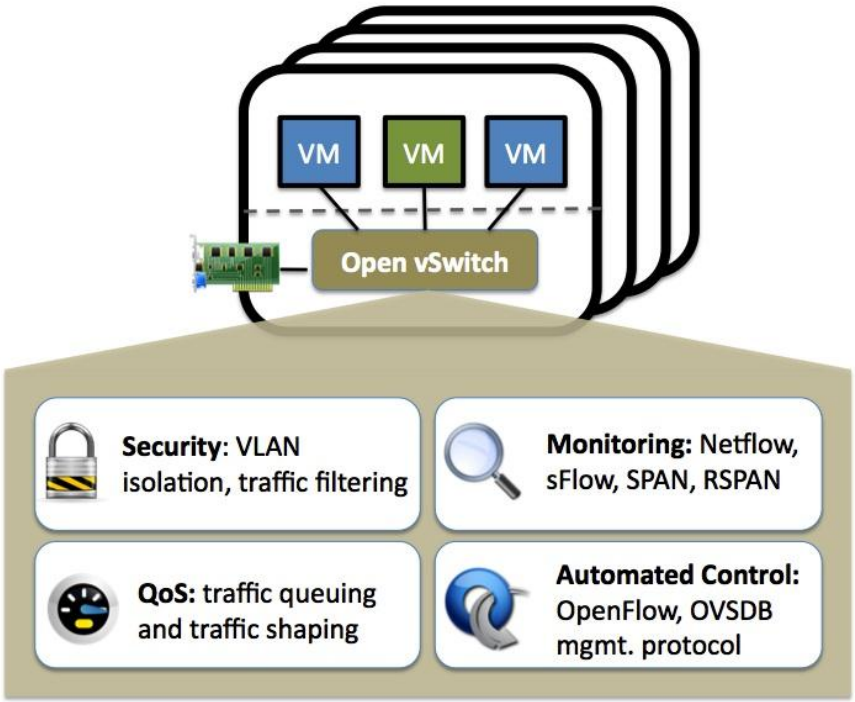
OF Controllers



- Several controllers out there (NOX, POX,Trema…)
- Floodlight from Big Switch. Apache license

cloudstack

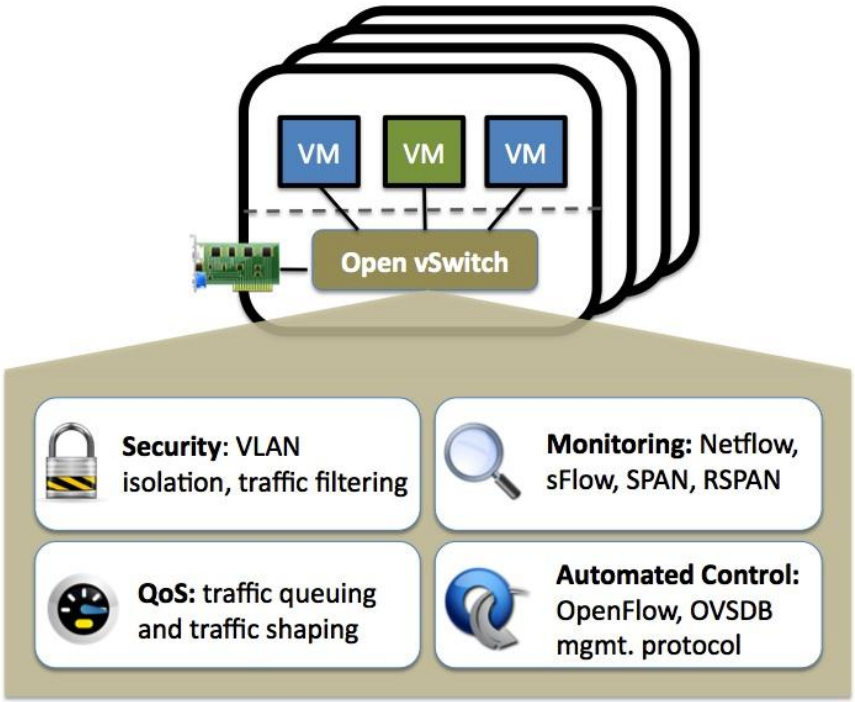
OpenVSwitch



- *"Open vSwitch is a production quality, multilayer virtual switch licensed under the open source Apache 2.0 license. It is designed to enable the massive network automation through programmatic extension…"*

cloudstack

OpenVSwitch

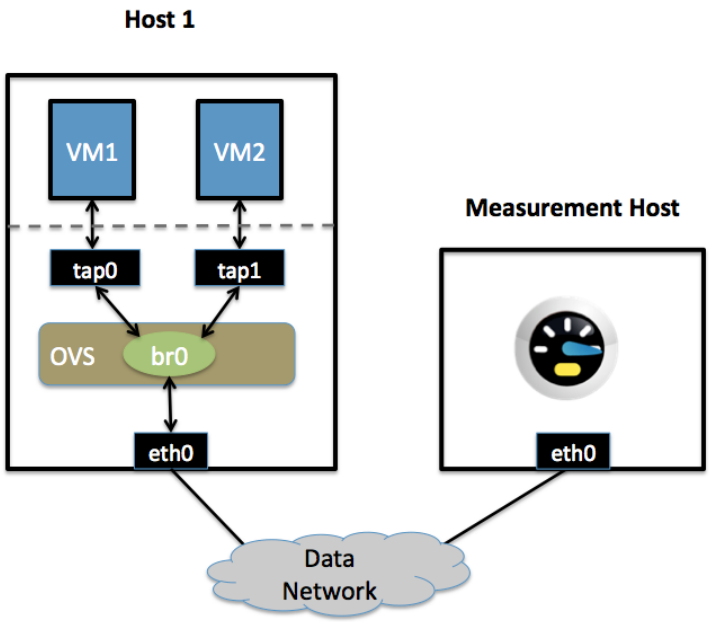


- Default bridge in XenServer and XCP
- Supported in Xen but not integrated in toolstack
- Enables:
  - VLAN tagging
  - Rate limiting
  - GRE tunnels
  - OpenFlow controller
  - …
- High Performance (http://networkheresy.com/category/open-vswitch)

cloudstack

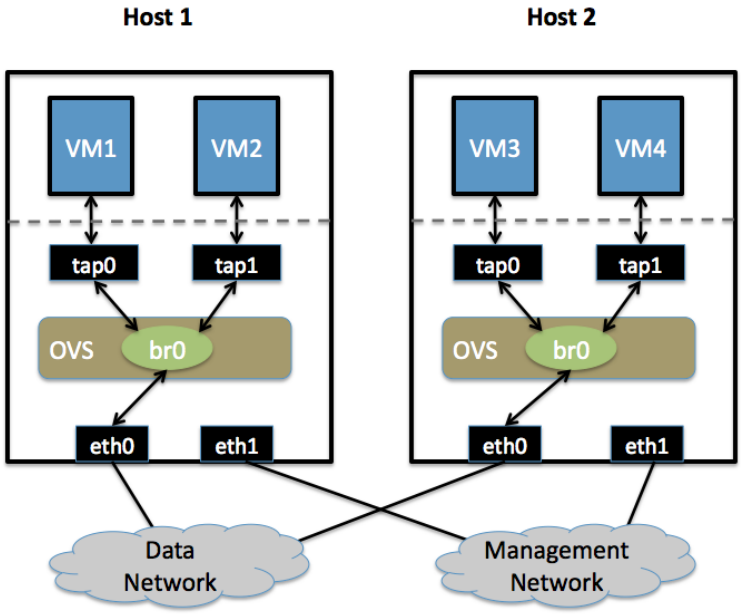
e.g OVS rate limiting

- Can enforce QoS with rate limiting controls
- *ovs-vsctl set Interface tap0 ingress_policing_rate=1000*
- *ovs-vsctl set Interface tap0 ingress_policing_burst=100*



cloudstack

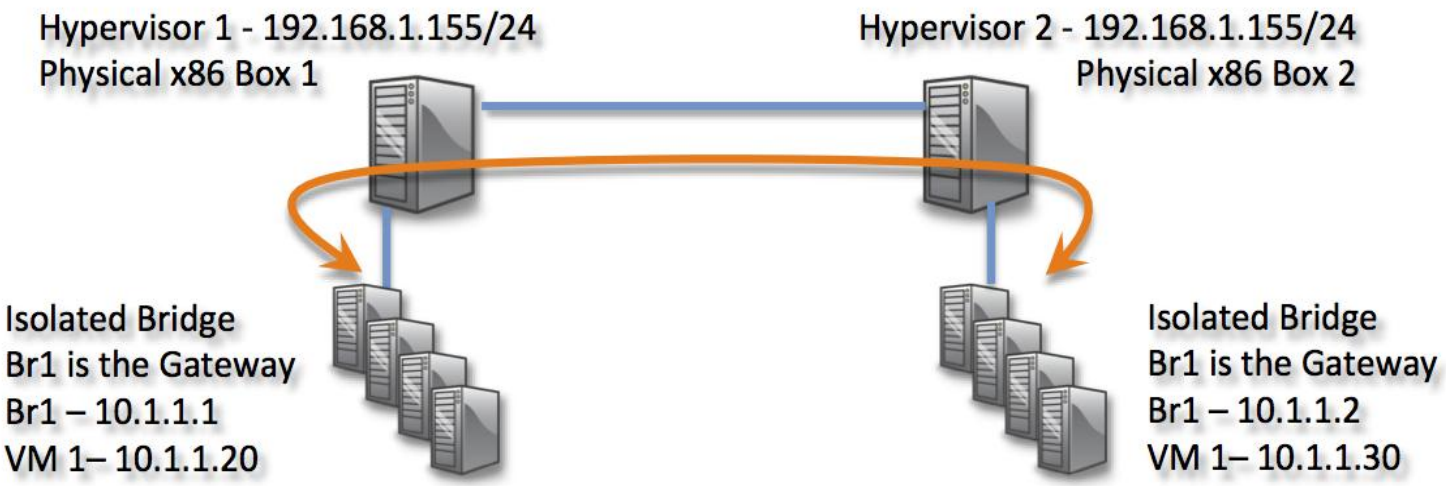
e.g OVS VLAN tagging



- *ovs-vsctl add-br br0*
- *ovs-vsctl add-port br0 eth0*
- *ovs-vsctl add-port br0 tap0 tag=1*
- *ovs-vsctl add-port br0 tap1 tag=2*
- Complement on host2…

cloudstack

e.g OVS and GRE tunnels



- No Cookbook on OVS page
- ovs-vsctl add-port br1 gre1 -- set interface gre1 type=gre options:remote_ip=192.168.1.152

cloudstack

OVS and Openflow

- Point OVS switches to an OF controller:

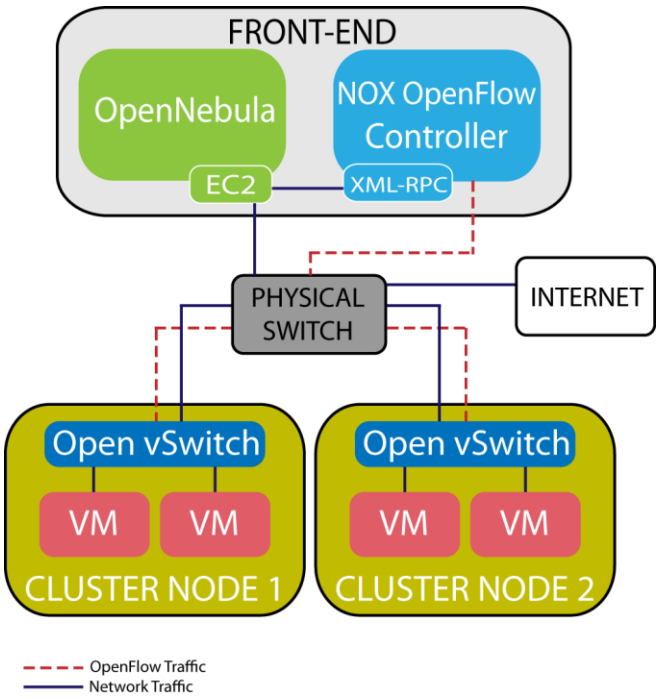
  $ovs_vsctl set-controller br0 tcp 192.168.1.33:6633

- Install rules on switch
  - Proactively (before any packet flows)
  - Reactively (unknown packets forwarded to controller, who pushes flow mod on switch, then operates at line rate)
- **Can do SDN with OpenFLOW but also with straight up OVS and managing mappings/rules in CloudStack db.**

cloudstack

OpenNebula

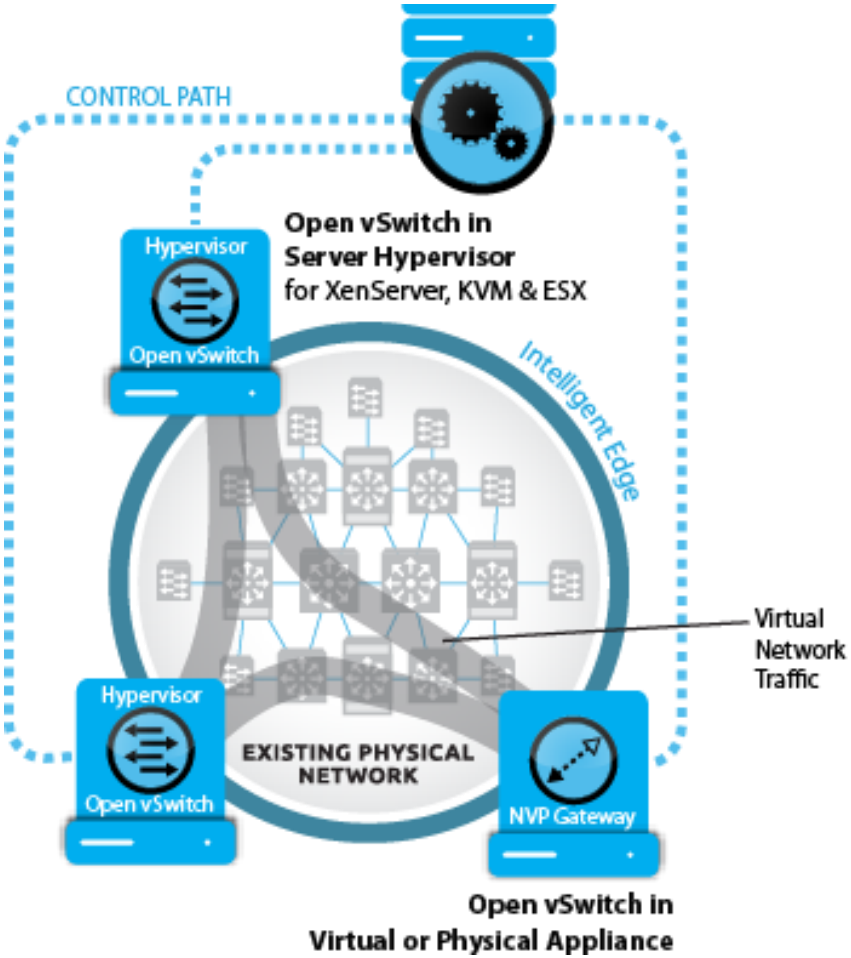
OpenNebula.org



- Supports VLAN tagging and rate limiting through “hooks” that call ovs_vsctl
- Scripts executed on an hypervisor before a VM is launched
- Potentially also executed after VM shutdown for cleanup
- Also supports OpenFlow

cloudstack

CloudStack Nicira Support



- https://cwiki.apache.org/confluence/display/CLOUDSTACK/Feature+Nicira+NVP+integration
- By Hugo Trippaers, Schuberg Philis

API key to customization of the network

Network

- createNetworkOffering
- updateNetworkOffering
- deleteNetworkOffering
- listNetworkOfferings
- createNetwork
- deleteNetwork (A)
- listNetworks
- restartNetwork (A)
- updateNetwork (A)
- createPhysicalNetwork (A)
- deletePhysicalNetwork (A)
- listPhysicalNetworks
- updatePhysicalNetwork (A)
- listSupportedNetworkServices
- addNetworkServiceProvider (A)
- deleteNetworkServiceProvider (A)
- listNetworkServiceProviders
- updateNetworkServiceProvider (A)
- createStorageNetworkIpRange (A)
- deleteStorageNetworkIpRange (A)
- listStorageNetworkIpRange
- updateStorageNetworkIpRange (A)
- addNetworkDevice
- listNetworkDevice
- deleteNetworkDevice
- listF5LoadBalancerNetworks
- listSrxFirewallNetworks
- listNetscalerLoadBalancerNetworks

Load Balancer

- createLoadBalancerRule (A)
- deleteLoadBalancerRule (A)
- removeFromLoadBalancerRule (A)
- assignToLoadBalancerRule (A)
- createLBStickinessPolicy (A)
- deleteLBStickinessPolicy (A)
- listLoadBalancerRules
- listLBStickiness
- listLoadBalance
- updateLoadBal
- addF5LoadBala
- configureF5Loa
- deleteF5LoadB
- listF5LoadBalar
- addNetscalerLo
- deleteNetscaler
- configureNetsc
- listNetscalerLo

Firewall

- listPortForwardingRules
- createPortForwardingRule (A)
- deletePortForwardingRule (A)
- createFirewallRule (A)
- deleteFirewallRule (A)
- listFirewallRules
- addSrxFirewall (A)
- deleteSrxFirewall (A)
- configureSrxFirewall (A
- listSrxFirewalls

Router

- startRouter (A)
- rebootRouter (A)
- stopRouter (A)

NAT

- enableStaticNat
- createIpForwardingRule (A)
- deleteIpForwardingRule (A)
- listIpForwardingRules

VPN

- createRemoteAccessVpn (A)
- deleteRemoteAccessVpn (A)
- listRemoteAccessVpns

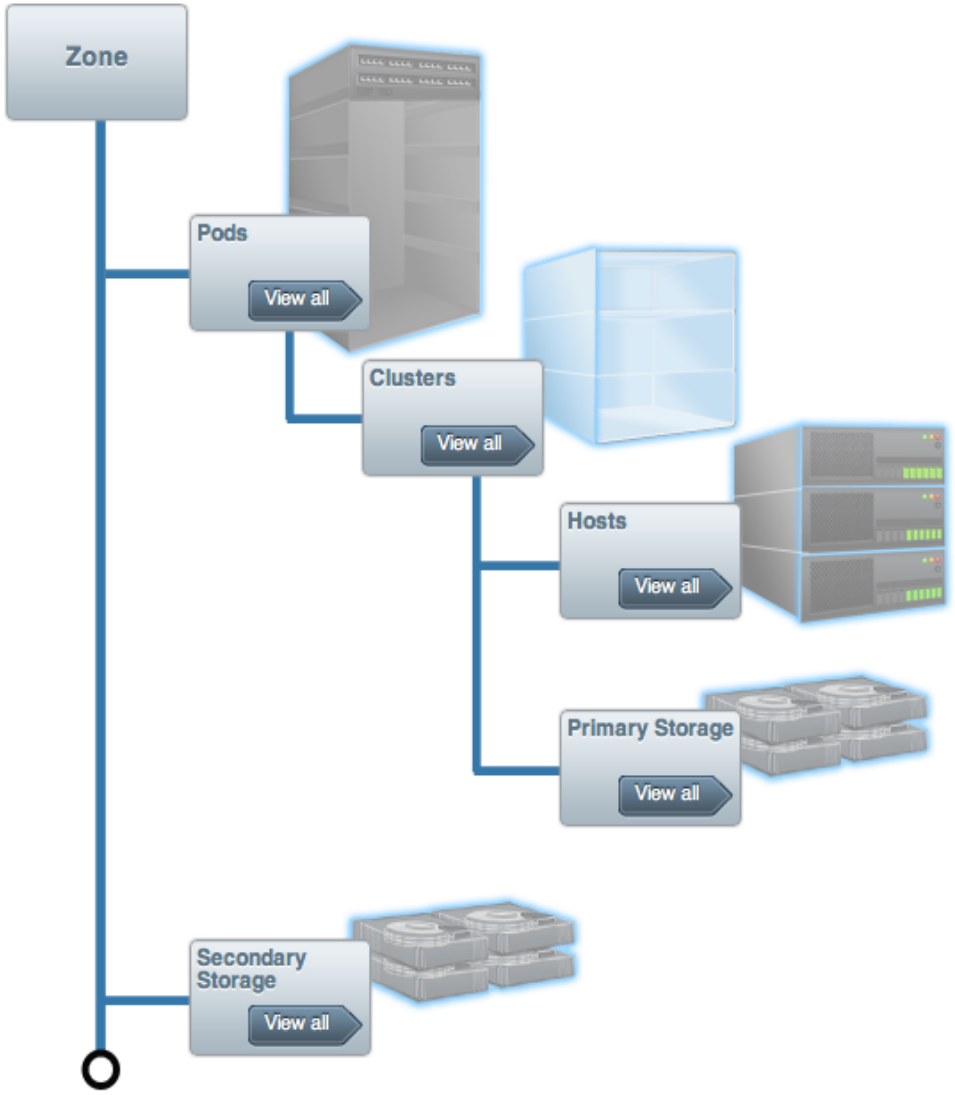
VLAN

- createVlanIpRange
- deleteVlanIpRange
- listVlanIpRanges

- You dream it, CloudStack orchestrates it ☺

cloudstack

Terminology



Zone: Availability zone, aka Regions. Could be worldwide. Different data centers

Pods: Racks or aisles in a data center

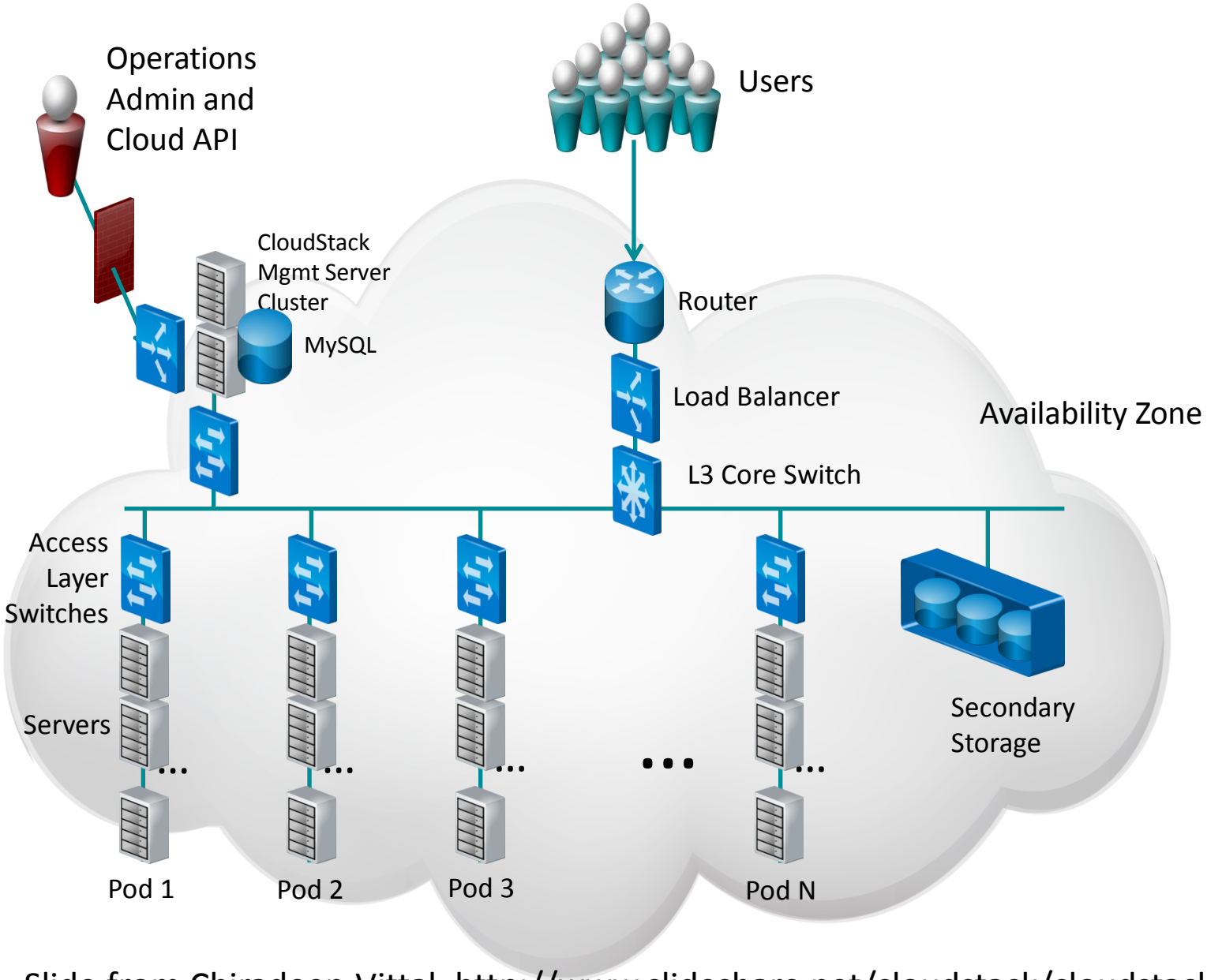
Clusters: Group of machines with a common type of Hypervisor

Host: A Single server

Primary Storage: Shared storage across a cluster

Secondary Storage: Shared storage in a single Zone

Physical Network

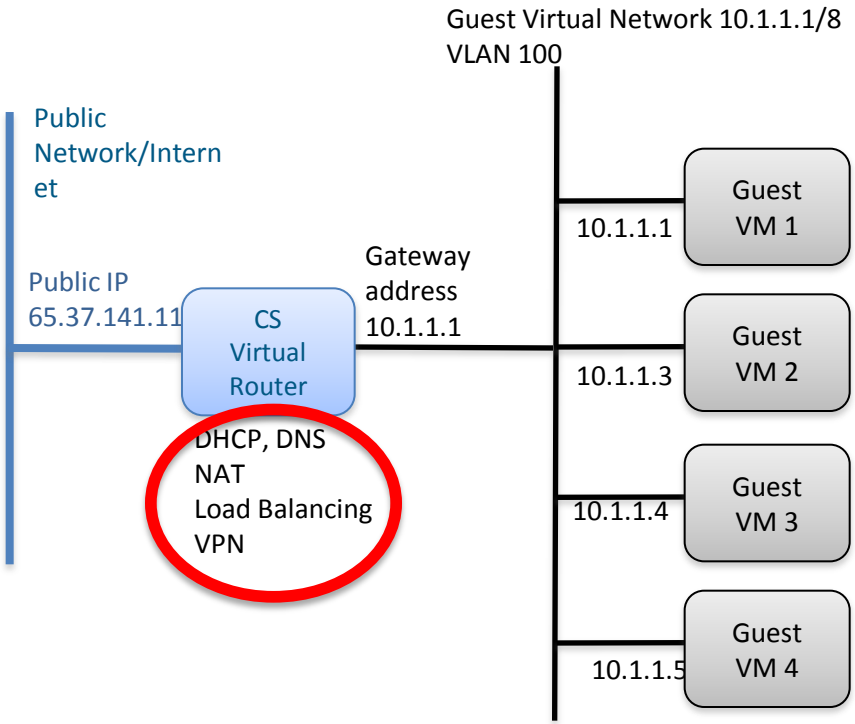


Slide from Chiradeep Vittal, http://www.slideshare.net/cloudstack/cloudstack-networking

cloudstack

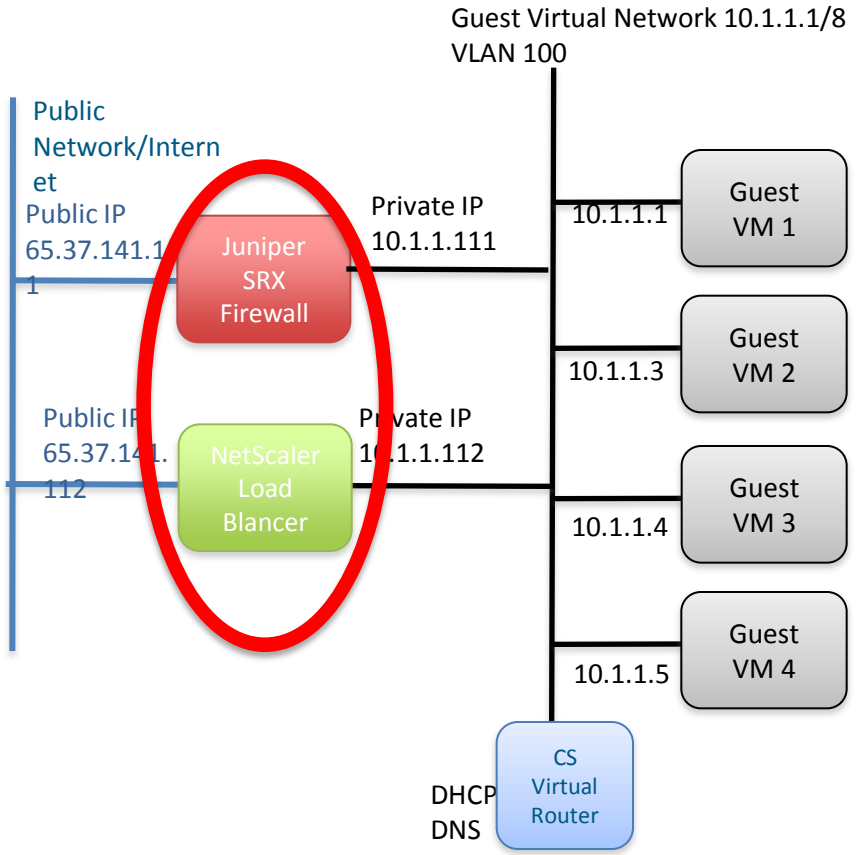
Layer-2 Guest Virtual Network
1 VLAN per guest network

CS Virtual Router provides Network Services



External Devices provide Network Services

Network Hardware exposing API can be controlled



Slide from Chiradeep Vittal, http://www.slideshare.net/cloudstack/cloudstack-networking

cloudstack

Opportunity for Xen

- Opportunity to create highly specialized networking services appliances using
  - OpenMirage VMs
  - HalVM
- See talks in Monday's session

cloudstack

Networking challenges in a private Cloud

- Multi-tenants on hypervisors => isolation between guest networks
- VLANs in the datacenter is hard and **limit at 4096 VLANs.**
- Hardware switches may not do it very well or have a lower limit

cloudstack

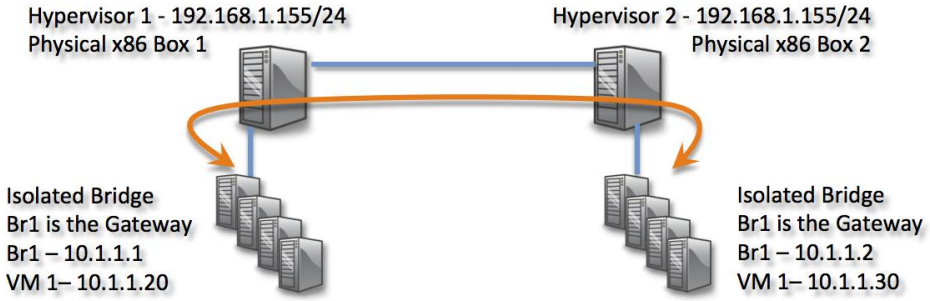
Networking trend

- Move to software switches
- Move to L3 isolation
- Use tunnels between OVS (GRE tech preview)
- **Program the network through API**
- Encapsulation virtualizes the network, between overlays on overlays on overlays..
- L3 on L2 on GRE on L3 on L2…
- Then you bring the WAN and you have:
- **L3 on L2 on GRE on L3 on L2 on GRE on L3 on L2 ….Euhhhh !!!**

cloudstack

Back of the enveloppe

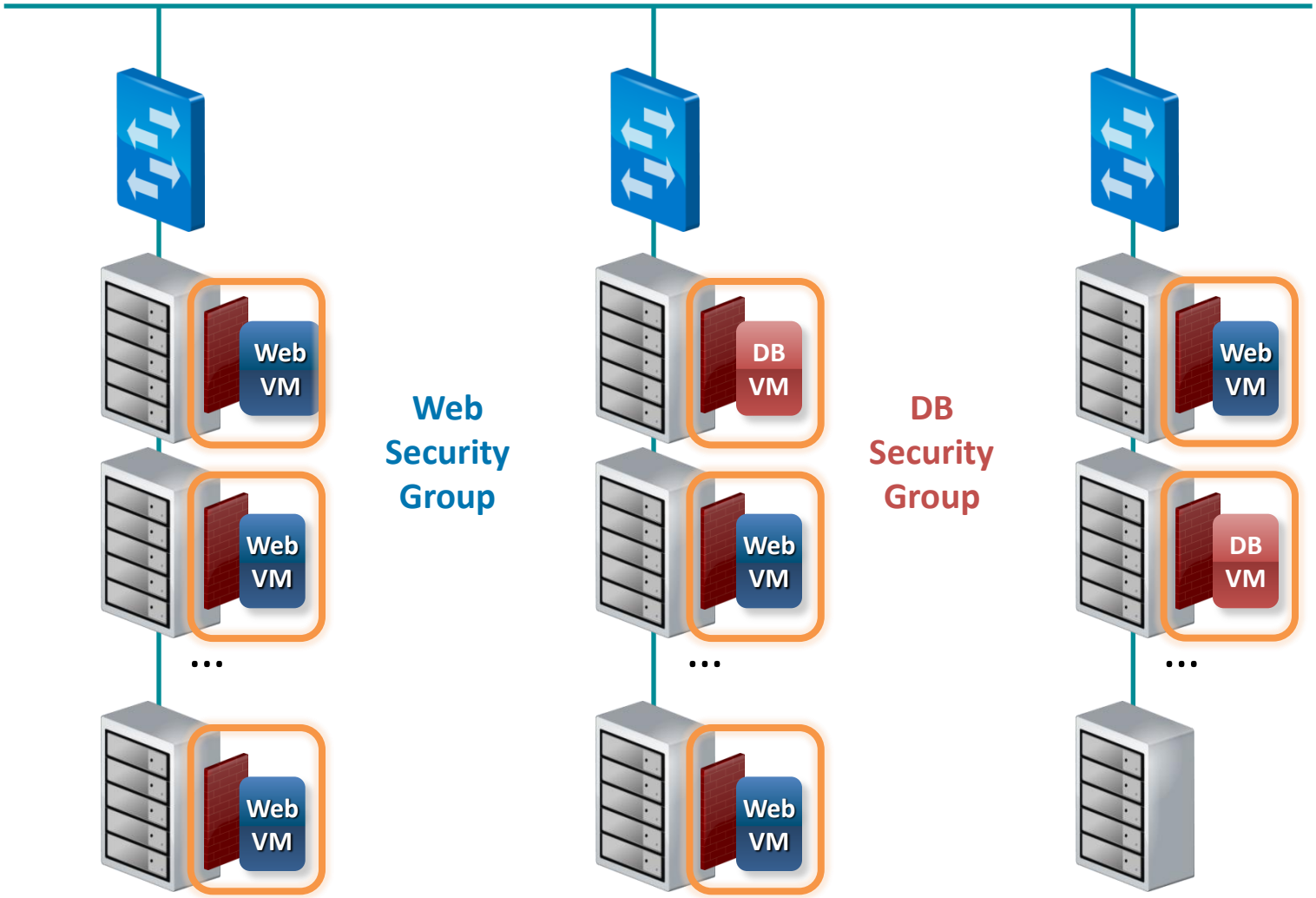
- ~10,000 hypervisors in your data center
- ~100,000 VMs
  - x10 or x100 if you use HalVM or Openmirage.org ☺
- (10,000*9,999)/2 tunnels for a full mesh
  - 50x10^6 tunnels to keep track of ?



cloudstack

Slide from Chiradeep Vittal

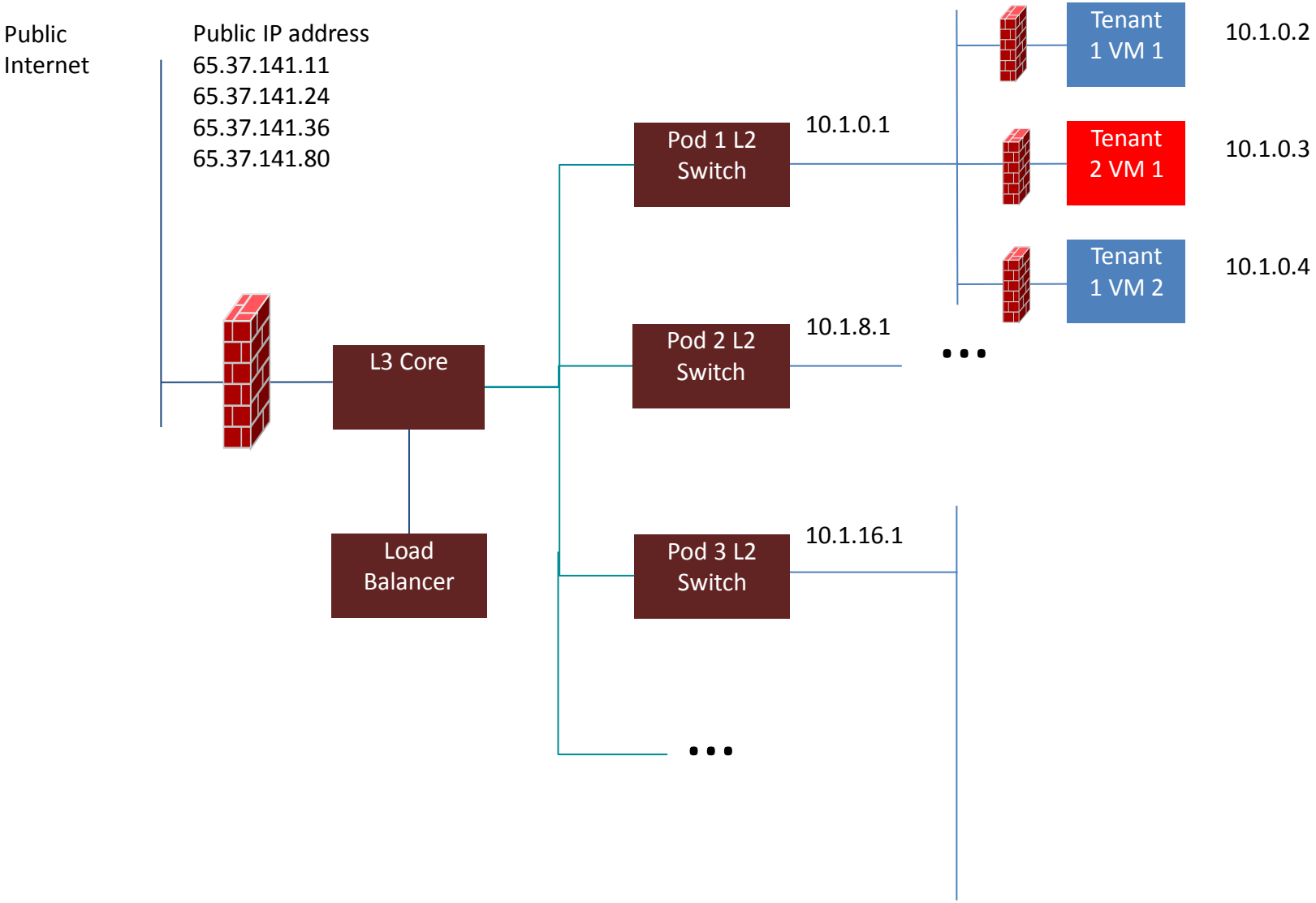
Layer 3 cloud networking



Ingress Rule: Allow VMs in Web Security Group access to VMs in DB Security Group on Port 3306

cloudstack

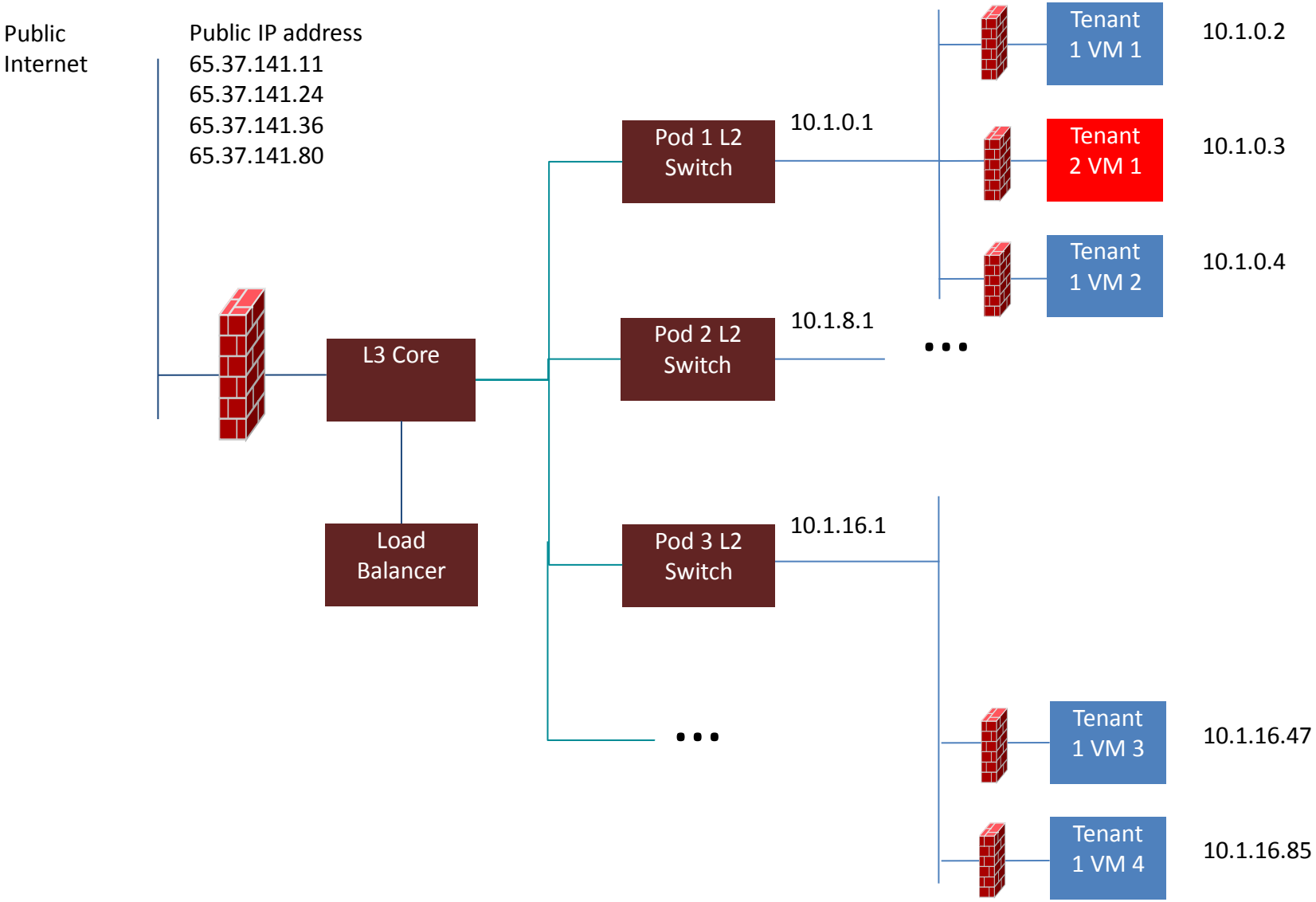
L3 isolation with distributed firewalls



Slide from Chiradeep Vittal

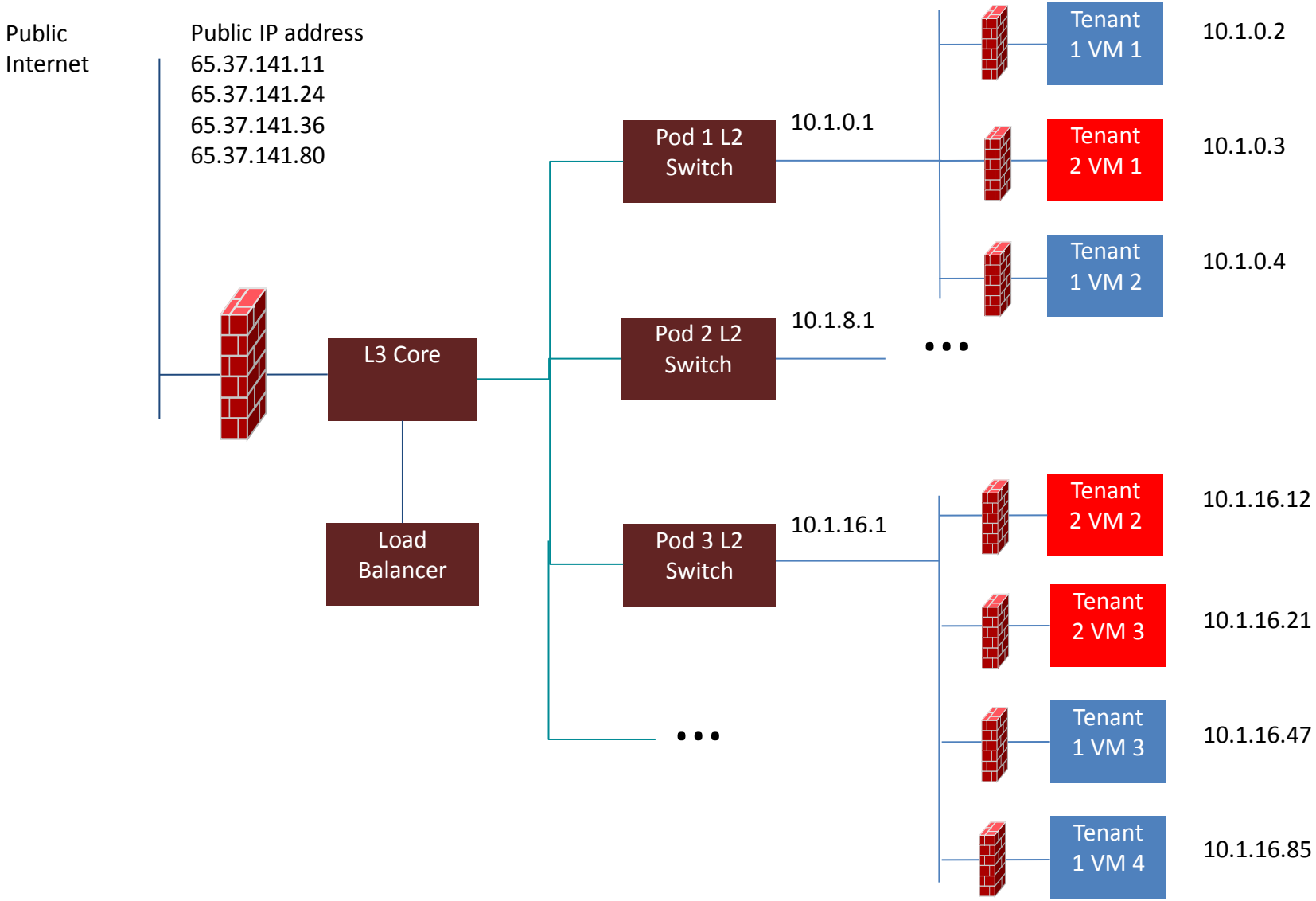
cloudstack

L3 isolation with distributed firewalls



Slide from Chiradeep Vittal

L3 isolation with distributed firewalls



Slide from Chiradeep Vittal

cloudstack

A Million Firewalls?

Slide from Chiradeep Vittal

cloudstack

Problem:

Manage the state of 100s of thousands of firewalls

Solution:

Well-known software scaling techniques

- Message queues
- Consistency tradeoffs
- Idempotent configuration & retries

CloudStack uses

- special purpose queues
- optimized for large security groups
- eventual consistency for rule updates

Slide from Chiradeep Vittal

cloudstack

Problem:

Firewall (iptables) rules explosion on the host firewall

Solution:

Use ipsets:

```
ipset –N web_sg  iptreemap
ipset –A web_sg 10.1.16.31
ipset –A web_sg 10.1.16.112
ipset –A web_sg 10.1.189.5
        …
ipset –A web_sg 10.21.9.77

-A FORWARD –p tcp –m tcp –dport 3060 –m  set –match-set web_sg src  -j ACCEPT
```

Slide from Chiradeep Vittal

cloudstack

Conclusions

- Programmable networking is here
- Software switches are key enabler to network virtualization
- Opens the door for scalable, on-demand, ephemeral networks
- OVS is the default switch in Xen, and supported in XenServer and XenCP.
- CloudStack implements highly scalable network structures and leverages OVS capabilities

cloudstack

Participating in CloudStack

- Apache incubator project
- http://www.cloudstack.org
- #cloudstack on irc.freenode.net
- @CloudStack on Twitter
- http://cloudstack.org/discuss/mailing-lists.html

Welcoming contributions and feedback, Join the fun !