Cloudcasting

A New Architecture for Cloud Centric Networks

Richard Li, Kiran Makhijani, Lin Han

Huawei Technologies

America Research Center

Santa Clara, CA, USA

Agenda

- **Motivation**
  - Network virtualization overview
  - Gaps in current approaches
- **Cloudcasting**
  - Architecture & Operation
  - Deployment scenarios
- **Analysis**
  - Comparison
  - Benefits
  - Implementation
- **Conclusion**

Paradigm Shift - Traditional to Cloud-Oriented Datacenters

- Virtualization has changed the behavior of datacenters infrastructures
  1. **Configuration**
  - Static configuration and dedicated resources do not work

  ✓To be dynamic and share resources across the system

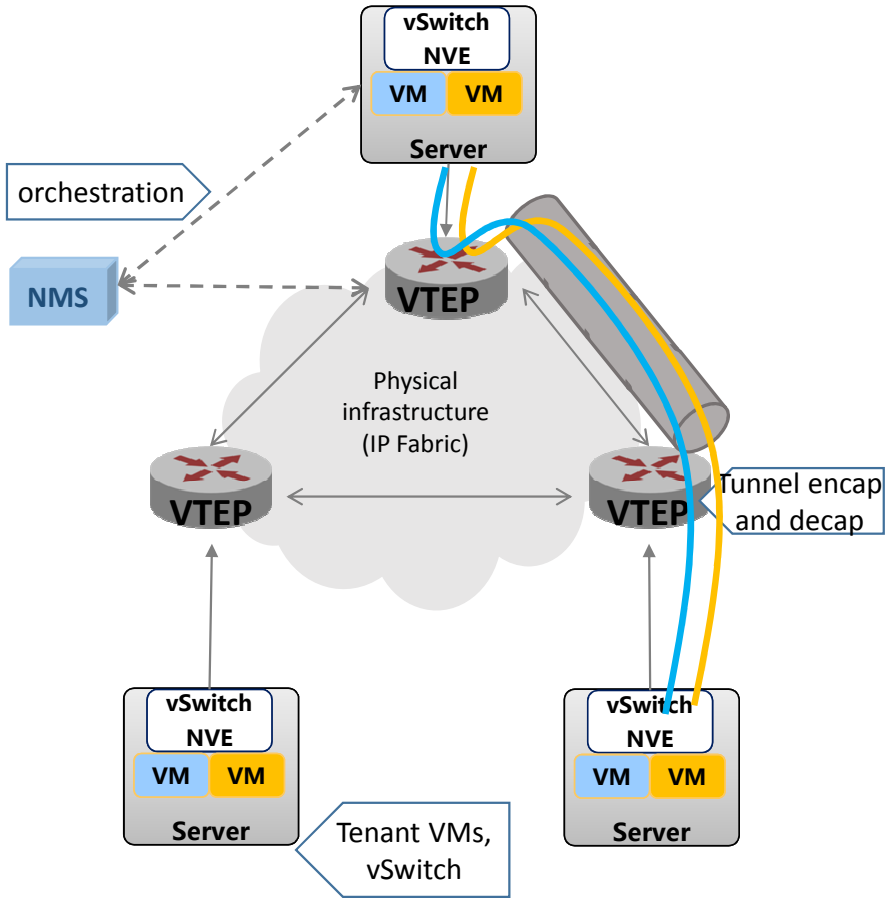
  2. Applications
  - Servers/appliances are not at fixed network locations

  ✓Are location independent and distributed

  3. Scale of Resources
  - Data center resource growth by scaling up

  ✓Instead grow and shrink horizontally on-demand for optimal utilization
- Modern Datacenters are virtualized, elastic and massive scale

Network Virtualization - Overview



Ability to create logical, tenant networks that are decoupled from the underlying physical infrastructure (Substrate network)

VTEP (Virtual Tunnel Endpoint)

- Network virtualization edge with address in Infrastructure
- Maintains mapping of VMs/devices in a virtual network to remote VTEP
- Encapsulation and decapsulation functions

NV Solution Readiness for cloud based data centers?

- Dynamic interconnection of appliances in tenant networks
- Suitable scale, flexibility…

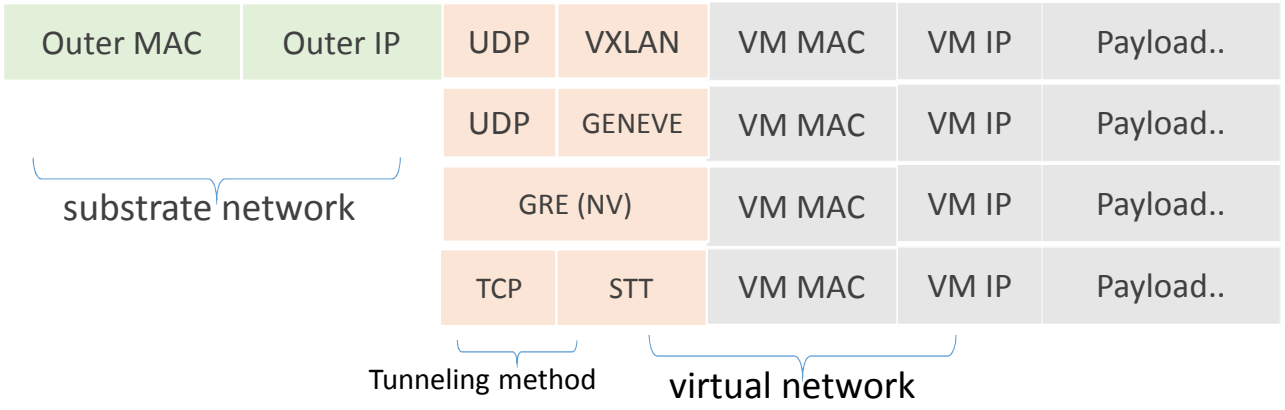
Network Virtualization in Data plane

Virtualization through overlays .e.g.,

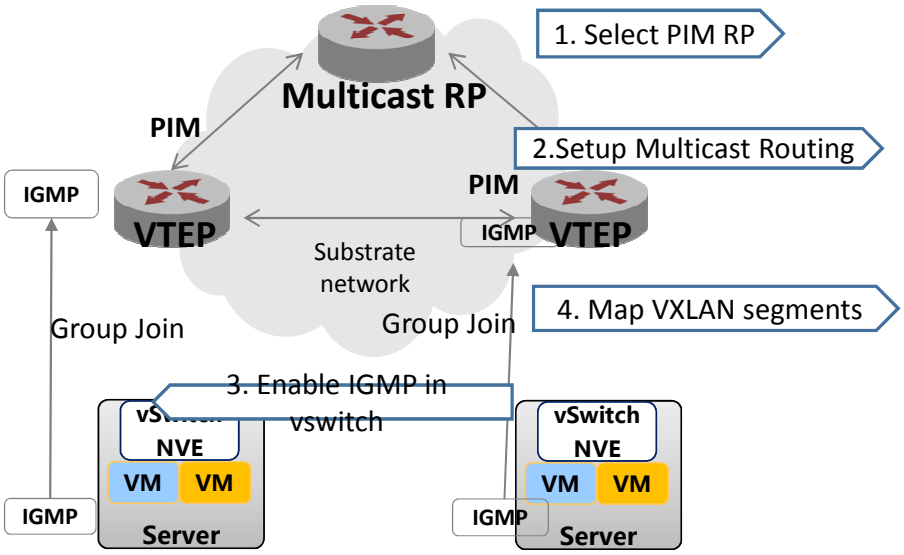
- Virtual Extensible LAN – VXLAN
- Network virtualization using GRE– NVGRE
- And now GENEVE
- OTV, SPB and TRILL (for Inter DC)

Focus & Benefits

- Virtual network connectivity extends over Layer 3/IP fabric
- Support isolation between tenants and workload migration
- Simple hardware friendly encapsulation



Network Virtualization Control plane - Multicast




```
feature ospf
feature pim
router ospf 1
  router-id 100.100.100.1
  ip pim rp-address 10.1.1.1
  group-list 224.0.0.0/4 bidir
```

```
interface loopback0
 ip address 100.100.100.1/32
 ip router ospf 1 area 0.0.0.0
   ip pim sparse-mode
```

```
feature nv overlay
feature vn-segment-vlan-based
interface e1/1
   switchport
   switchport access vlan 10
   no shutdown
interface nve1
   source-interface loopback0
   member vni 10000 mcast-group
230.1.1.1
   vlan 10
   vn-segment 10000
```

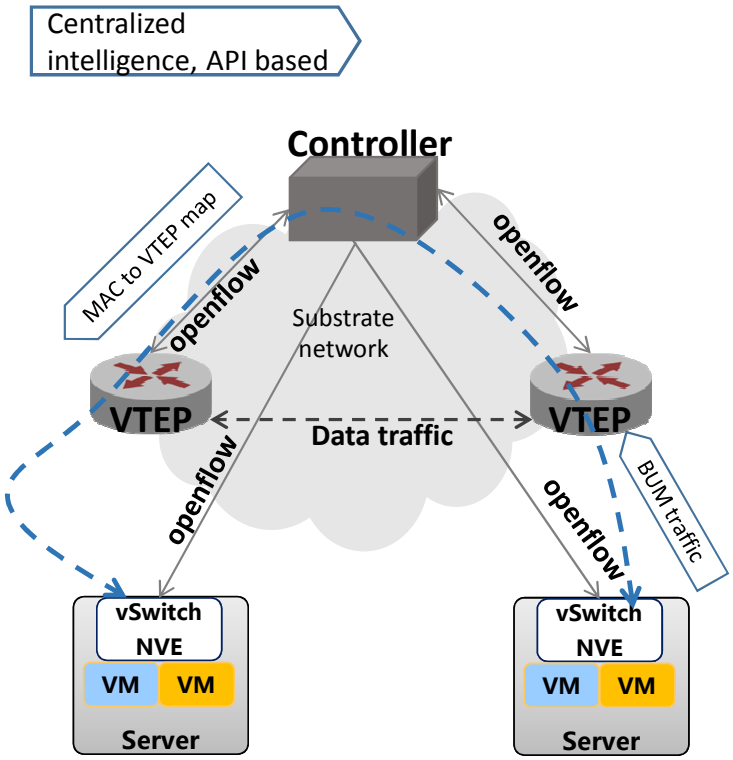
How It Works

- IP Multicast distributes VNI segment to MAC mapping through flooding on multicast group.
- VM discovery through ARP broadcast over VXLAN segment-specific multicast destination.
- Remembers remote MAC-to-VTEP mapping

Challenges

- IP Multicast Routing
- Still performs flooding of traffic
- Overheads on substrate network as tenants Networks change – More control plane traffic
- Not suitable for inter-data center

Network Virtualization Control plane Centralized SDN



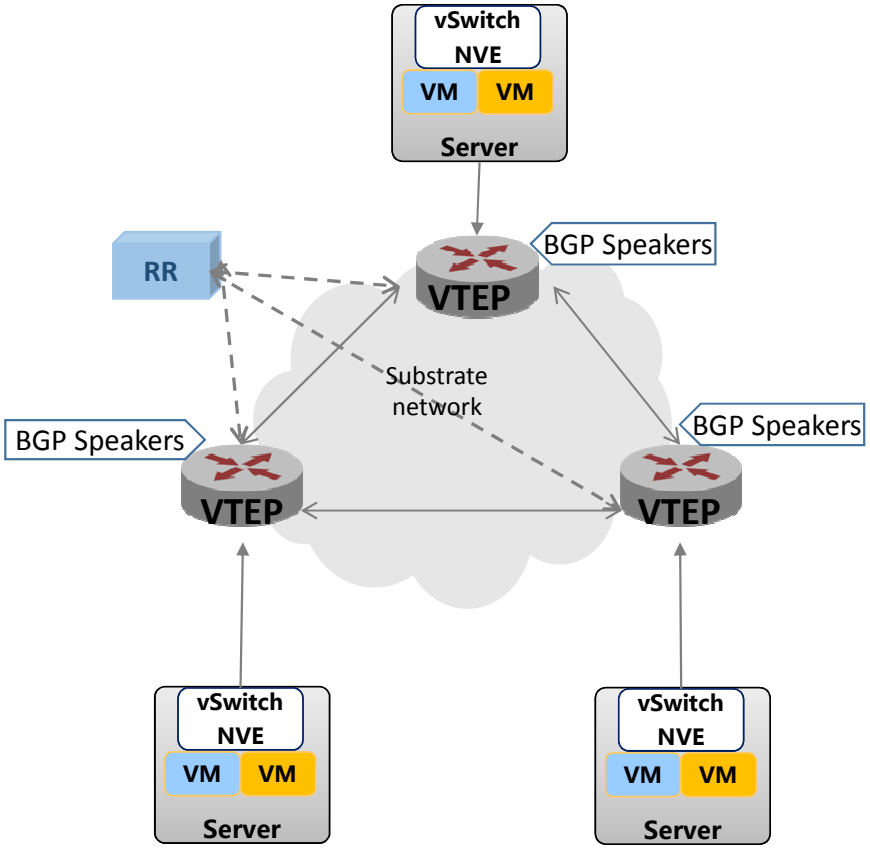
How It Works

- Centralized controller with full view of the network
- VM discovery through ARP unicast to controller or default gateways
- Centralization takes care of unknown floods traffic

Challenges

- Complexity of network state moved to the controller.
- Slow network changes - always go through the controller. Not ideal for. performance under high rate of change.
- No standard auto-discovery; proprietary controllers.
- Multiple administrative domain challenges

Network Virtualization Control plane - MP-BGP



How It Works

- VM Routes distribution through MP-BGP protocol
- Virtual Networks discovery via route reflectors
- No unknown flooding
- BGP advantages - Robust and scales well

Challenges

- Complex Network design
  - IBGP (simple, full-mesh), EBGP (multiple AS, one per tenant)
- Mandates substrate network to be BGP
- Non-trivial configurations
  - Dependency between substrate and virtual network config
  - VPN style objects; linear growth with scale

Network Virtualization – Summarizing Challenges

A. Connectivity

- Should support multiple tenants or user networks
- Site/location independent - common framework for local and remote

✓Segmentation through network overlays

- Not uniform. various approaches. no single converged solution that worked for both.

B. Control Plane

- Should be infrastructure independent
- Should support route distribution and virtual network discovery at scale

- Un-coordinated, vendor specific, SDN controllers, multicast, BGP.
- Not abstract enough. Embeds into infrastructure protocol.
- Not elastic. Virtual network changes impact infrastructure

C. Data plane

- Many encapsulations in network virtualization
- Scale of virtual networks

✓Common theme – overlays (IP in IP, L2 in IP).

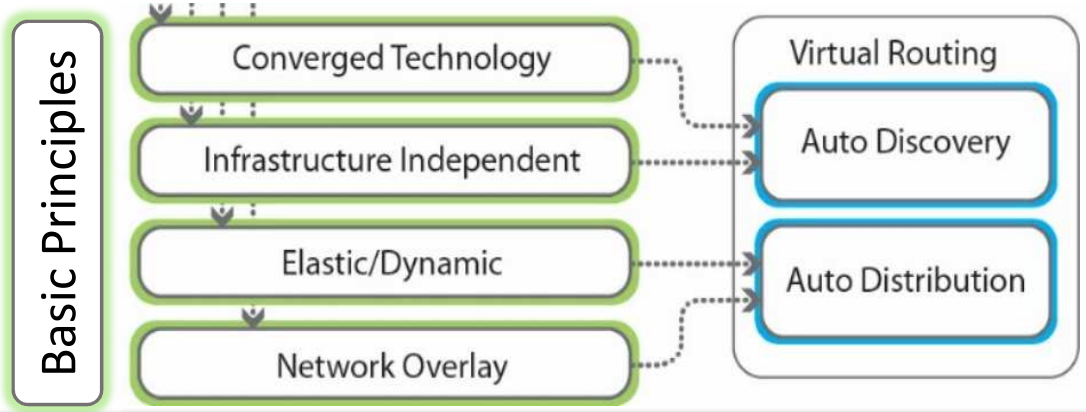
✓16 million VXLANs enough for now but with cloud based services, IoT. nice to have higher scale.

Agenda

- **Motivation**
  - Network virtualization overview
  - Gaps in current approaches
- **Cloudcasting**
  - Architecture & Operation
  - Deployment scenarios
- **Analysis**
  - Comparison
  - Benefits
  - Implementation
- **Conclusion**

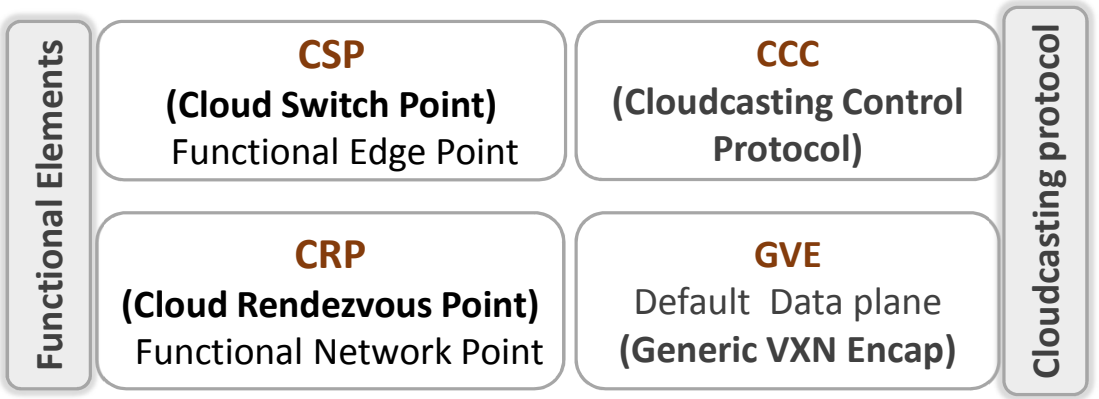
Cloudcasting Architecture Solution

- Eliminate complex network design
- A native control plane for virtual networks
- Minimize configurations

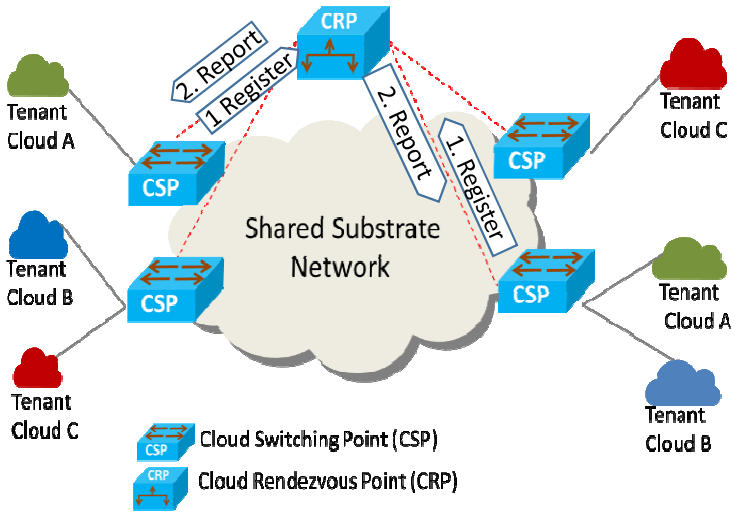


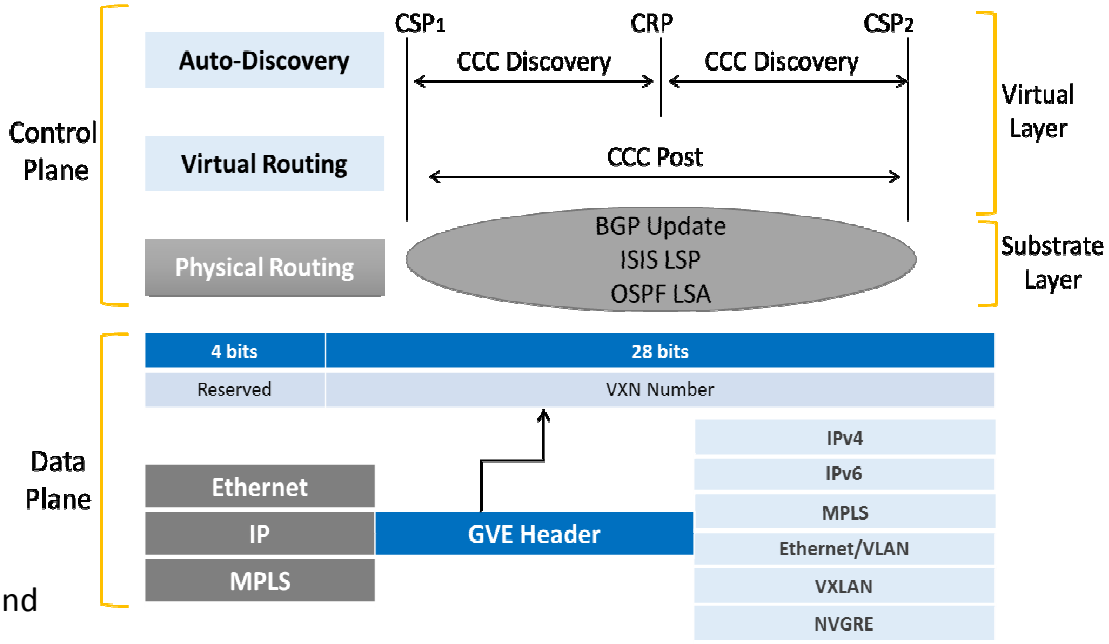
Cloudcasting Framework

1. Signaling – Register, Report, Post
2. CSP, CRP - protocol speakers
3. GVE - Higher scale data plane



Cloudcasting Control Protocol Primitives





CRP (Cloud Rendezvous Point)

- A single logical entity that stores information about CSP and their VXN participation
- Generates Reports towards CSP for discovery of VXN

CSP (Cloud Switch Point)

- an edge of a virtual network
- participates in auto-discovery by sending Register to CRP
- Route-distribution through Post updates between CSPs

VXN (Virtual Extensible Network)

- An identifier for a virtual network (28-bit)

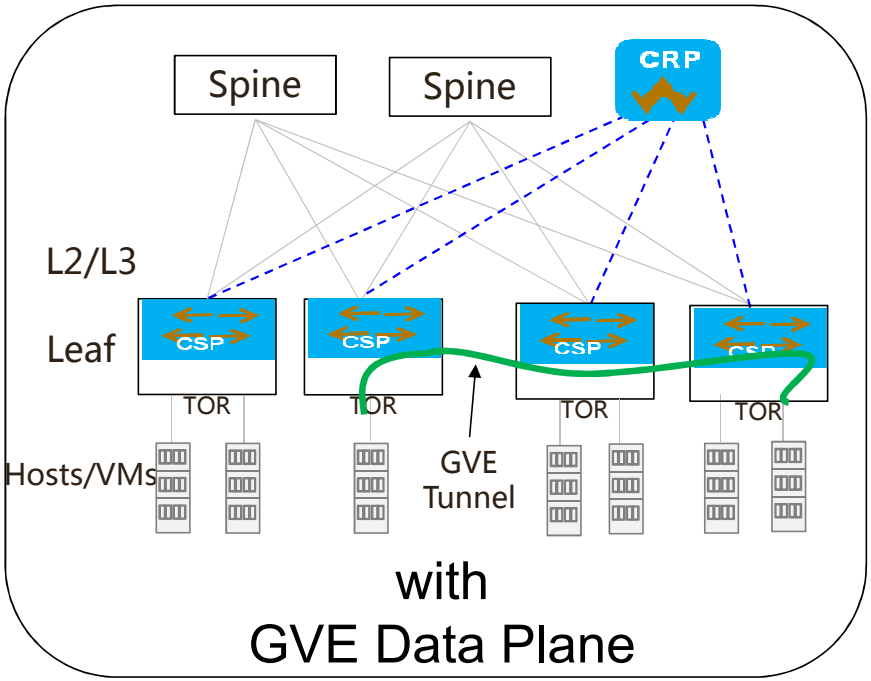
GVE (Generic VXN Encapsulation)

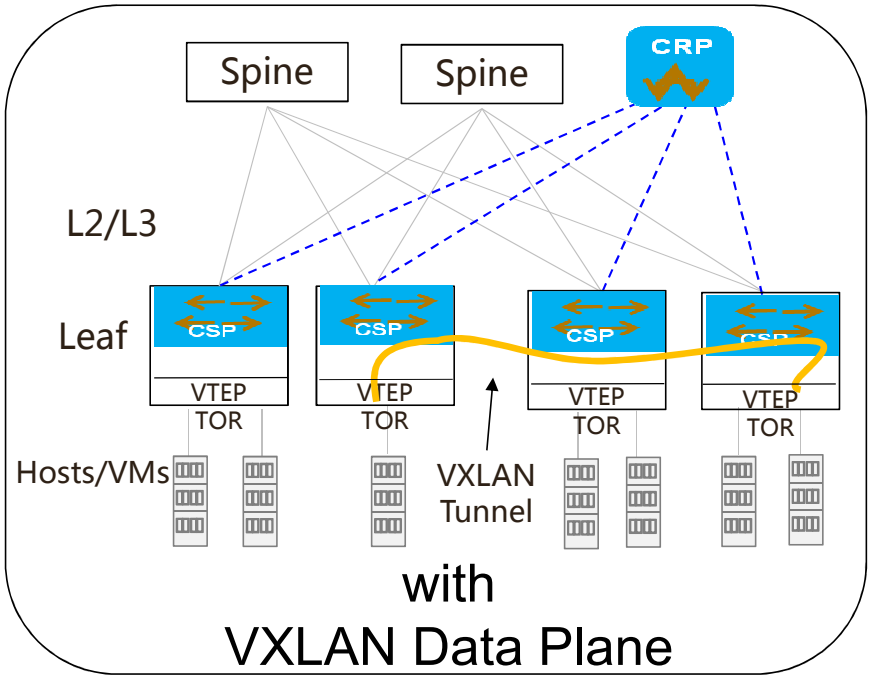
- Default 32-bit Data plane, can carry any overlay format
- Flexibility to overlay on MAC, IP or MPLS header

Cloudcasting Deployment Scenarios

- Multi-Site, Multi-Tenant Virtual Data Center
- Provide Cloud Services to Business VPN Customers
- Cloud Interconnect –public/public and public/private
- Provide Customized Virtual Networks for Applications
- OTT services through Virtual Network Operators

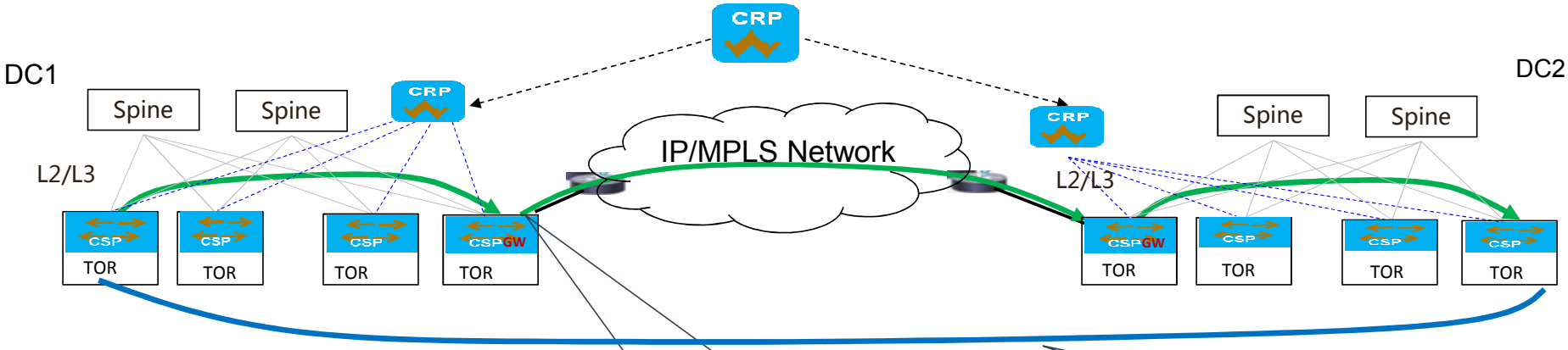
Multi-Tenant Data Center - Unified control plane





1. **Unified control plane support various data plane encapsulation**
2. **Virtual Network routes are not distributed everywhere in Infrastructure**

Data Center Interconnect



Converged Virtual Routing

Direct discovery and route distribution
As against Intra-DC VxLAN x Inter-DC EVPN or MP-BGP

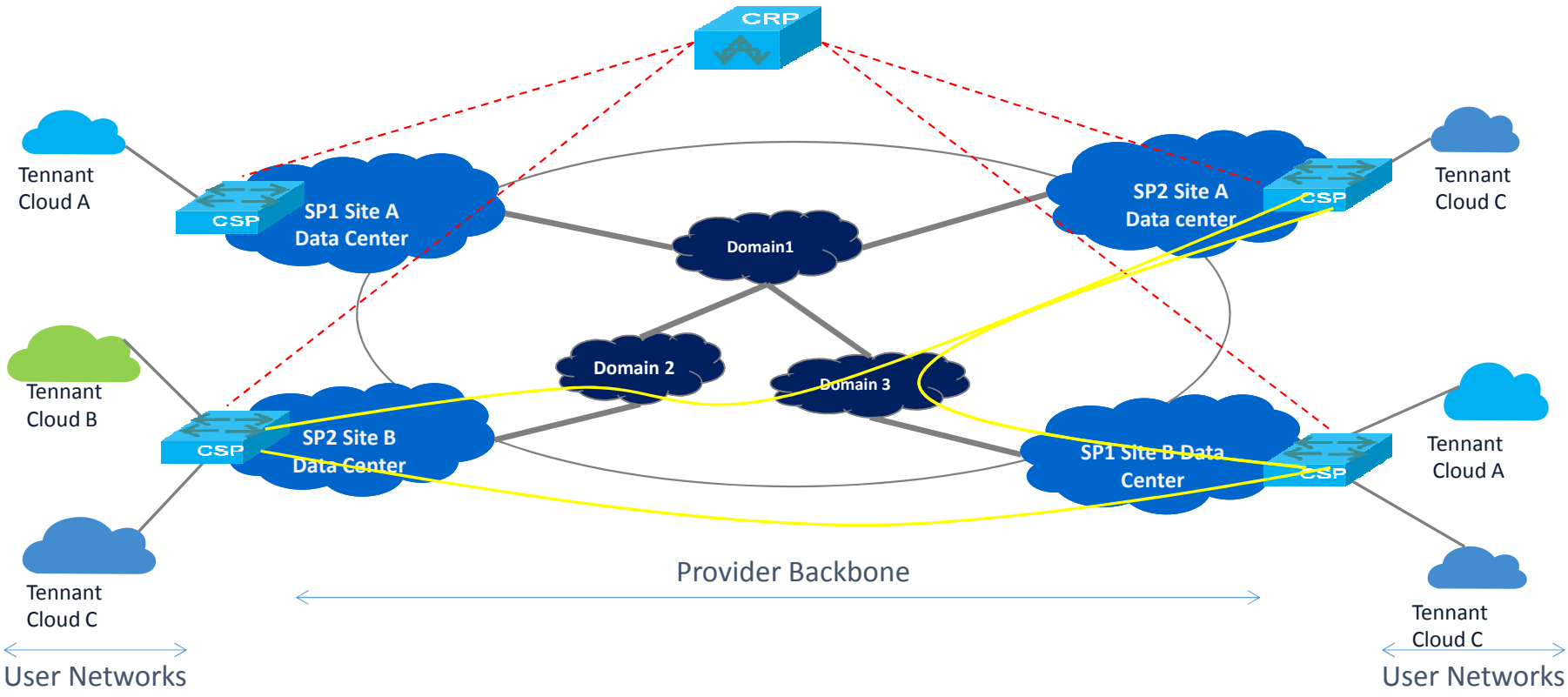
Multiple GVE Tunnels

CSP peering is limited to the local Cloudcast domain, so has the better scalability and manageability

Single GVE Tunnel

CSP peering is across the different Cloudcast domains

Cloudcasting Solution in Provider Space



Agenda

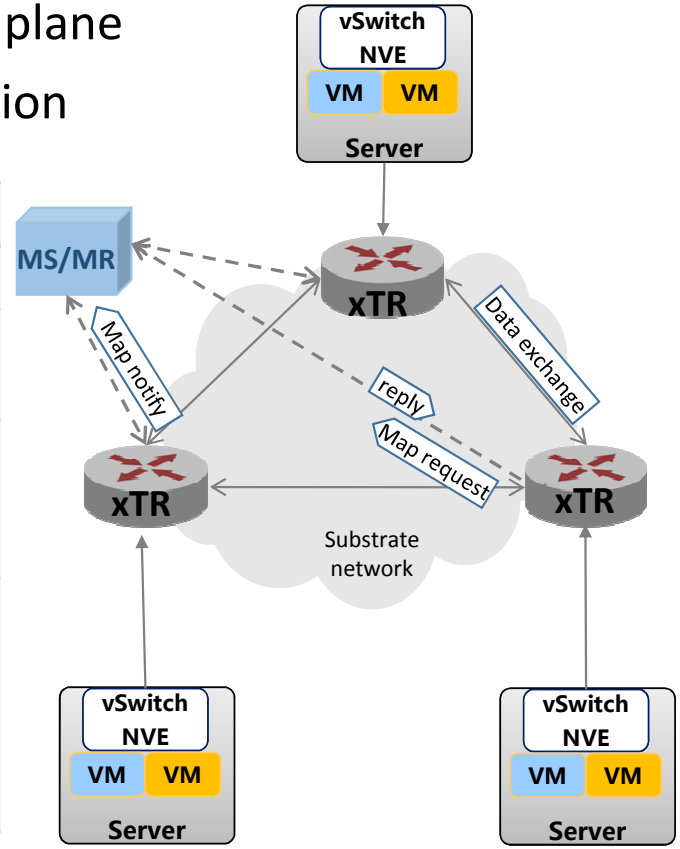
- **Motivation**
  - Network virtualization overview
  - Gaps in current approaches
- **Cloudcasting**
  - Architecture & Operation
  - Deployment scenarios
- **Analysis**
  - Comparison
  - Benefits
  - Implementation
- **Conclusion**

With Location/Identity Separation Protocol (LISP)

Only other protocol, that doesn't use infrastructure control plane

Uniform behavior for Inter- and Intra-data center virtualization

Function	LISP	Cloudcasting
Virtualization	• Instance Id (added later)	✓VXN
Address Family	• multiple address families (added later)	✓Same
Data plane	• Bigger header overhead • Reachability in data packets	✓ GVE is 32bits ✓ Rely on infrastructure.
Control plane	• Heavier signaling with mapping system • Data driven; Pull model (caching, aging)	✓Simpler, only VXNs are registered with CRP, not routes ✓Push model; easier stateless approach for networks changing dynamically.



Comparison with MP-BGP

Function	BGP	Cloudcasting
Virtualization	• VPN style	✓VXN
Address Family	✓ Through Multi-protocol	✓Through GVE
Data plane	• VXLAN, IP	✓ + more (NVGRE etc)
Control plane	Heavier signaling • route-distinguisher, route-target import/export • Setup VRFs, neighbors and AS • Provide community information	✓Simpler, ✓Neighbors are auto-discovered
Routing	• BGP peers receive and process much more than needed • Configuration proportional to scale of the system	✓Only per VXN routes distributed
VPN/DCI	• L3VPN and L2VPN need additional MPLS support. • Slower deployment	✓A Converged solution for all VPN, DCI and multi-tenant data center

Flavor of Configurations…

```
router bgp 100
  router-id 10.1.1.1
  address-family l2vpn evpn
     retain route-target all
```

```
vrf context vxlan-900001
   vni 900001
   rd auto
     route-target import 65535:101 evpn
     route-target export 65535:101 evpn
     route-target import 65535:101
     route-target export 65535:101
address-family ipv6 unicast
    route-target import 65535:101 evpn
    route-target export 65535:101 evpn
    route-target import 65535:101
    route-target export 65535:101
```

```
!
  neighbor 20.1.1.1 remote-as 200
     update-source loopback0
     ebgp-multihop 3
     address-family l2vpn evpn
        disable-peer-as-check
        send-community extended
        route-map permitall out
```

MP-BGP

```
! Configuration at router-B
!
router lisp
locator-set B
10.2.1.2

ipv4 itr map-resolver 10.0.14.2
ipv4 itr map-resolver 10.0.15.2
ipv4 etr
ipv4 etr map-server 10.0.14.2
```

```
Site A
Authentication-key site-a-passwd
eid-prefix 192.168.11.0/24
eid-prefix instance-id 1 192.168.14.0/24
eid-prefix instance-id 2 192.168.14.0/24
eid-prefix instance-id 3 192.168.14.0/24
```

```
eid-table vrf DeptB instance-id 1
database-mapping 192.168.16.0/24 locator-
set B
database-mapping 1:1:16::0/64 locator-set B
exit
```

LISP

```
router ccc
   router-id 20.1.1.1
   crp 100.100.100.100
!
```

```
vxn 100
 address-family vxlan
 network 192.168.16.0/24
 network 192.168.17.0/24
 exit
!
```

Cloudcasting

Blue blocks: one-time config

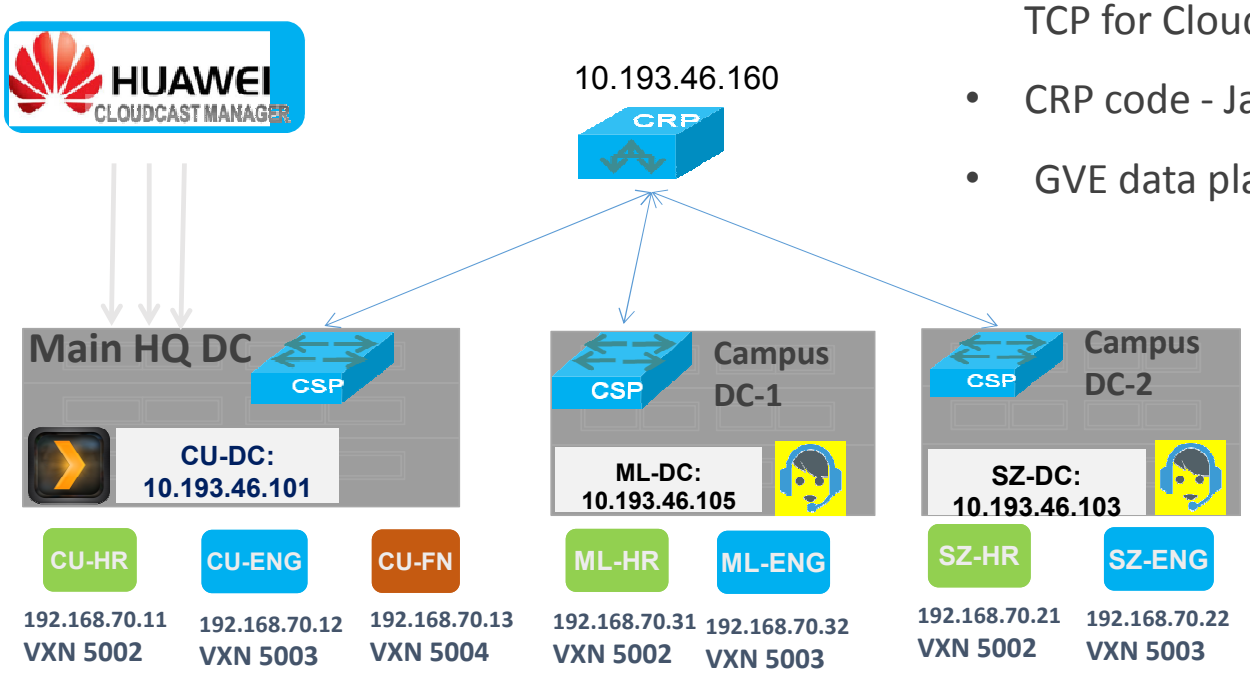
Black blocks
Linearly grow with VNs, prefixes

Simple Configuration => better agility

Cloudcasting satisfies cloud oriented network requirements

Converged	Same technology to build inter-, intra- data center, and VPNs
Elasticity	Any/heterogeneous protocols of the substrate network, Number of virtual networks, domains, routes within a user's network
Efficiency	No CSP distributes routes to other CSPs that they are not interested in
Distribution	change in the tenant networks can be announced immediately, no configuration changes Every time when a new CSP is added, it is only required to configure the newly added
Scale	Default GVE encapsulation support 256 million clouds

Implementation Details- Proof Of Concept



Development

- CSP Code - developed using quagga open-source.

  TCP for Cloudcasting control PDUs
- CRP code - Java-8/Neo4J for high scale database
- GVE data plane – process in user space

Demo

- Dashboard for VXN provisioning
- Video clients accessed video services from HQ

Closing Remarks

- Defined Converged Virtual Routing Concept
  - A pragmatic approach = best of prior knowledge + innovation
- Use cases
  - Multi-tenancy, Connection Clouds, VPN, Purpose built Networks
- Validated Cloudcasting vis a vis other solutions
- Next Steps – Beyond Reachability
  - Security, scale & policy framework
  - Design of client-server interface between virtual and physical network

Thank You