You make possible
Case study of SP customers running ACI based SDN for telecom datacenter

Abhishek Mande – Principle SE, SP
Sonu Khandelwal – TME, IBNG

BRKACI-3620
Cisco Webex Teams

Questions?
Use Cisco Webex Teams to chat with the speaker after the session

How
1. Find this session in the Cisco Events Mobile App
2. Click “Join the Discussion”
3. Install Webex Teams or go directly to the team space
4. Enter messages/questions in the team space
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
• Looking Ahead
• Conclusion
Session Objective

• This is a "Case Study" session, highlighting how large Telecom Service Providers used ACI to meet their requirements.

Initial assumption:
The audience already has knowledge of ACI concepts (Tenant, VRF, BD, EPG, L3Out, etc.)
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
• Looking Ahead
• Conclusion
Evolution of Telco Datacenter

Connectivity Service:
- EPC CUPS
- vRAN

Edge Computing:
- CDN, IOT, Localized AR, Analytics
- Analytics, IOT, Localized AR, Gaming, Public cloud hosting, CDN...

Control plane of CUPS gateways

Traditional IT apps

Managed Services:
- CPE / SD-WAN, FW IT Apps

Evolution of Telco Datacenter:
- Access / Cell sites
- Hub / Pre-aggregation
- Aggregation
- Regional DC
- Central DC
- Multiple public clouds
Agenda

• Evolution of Telco Datacenter

• Customer Profiles & Technical Requirements
  • Applications
  • Datacenter Fabric

• How ACI met the requirements

• Looking Ahead

• Conclusion
Customer Profiles

- 15+ Top operators worldwide
- Largest deployment of 60+ fabrics serving 300mn+ subscribers
- Largest fabric in telco DC 200+ leafs
- Fabric for 3G, 4G services (Physical & Virtual)
- Mix of Cisco and Non-Cisco 3G, 4G services
- Readiness of 5G in some advanced markets (Leading with NSA deployments)
## Telco Applications Requirement

<table>
<thead>
<tr>
<th>IMS</th>
<th>SGW &amp; PGW</th>
<th>Gi-LAN</th>
<th>CDN &amp; OTT Caching</th>
</tr>
</thead>
</table>
| • Faster Convergence  
• Consistent low latency  
• Multiple Hypervisor  
• IPv4 & IPv6 connectivity | • Virtualized deployment across multiple racks for control and data plane  
• Failover & Redundancy  
• Routing with fabric | • TCP Optimizers  
• Deep Packet Inspection (DPI)  
• CG-NAT  
• URL Filtering  
• Service Chaining with multiple nodes | • Caching managed by OTT Provider  
• Self created Media Content  
• Driving 100G  
• Migrating to 400G interfaces |

**Mobile Networking Monitoring**

• High performance data replication with scale

**SGW & PGW**

• IPv4 & IPv6 connectivity  
• Multiple hypervisor

**Gi-LAN**

• Faster Convergence  
• Consistent low latency  
• Multiple hypervisor

**Voice over Wifi**

**Assurance and insights**

• Proactive troubleshooting  
• Real time visibility  
• Analytics  
• Network assurance
Datacenter Fabric Requirement

- Fabric Automation
- Distributed DC
- Security & segmentation

- Policy based deployment
- Service Chaining
- Multi-hypervisor

- Fabric

- Operations
- North bound API
- Simplified Operation
- Troubleshooting

- Connectivity

- Availability
- Faster Convergence
- Carrier Grade Availability

Software Defined Network (SDN)
## Datacenter Fabric Protocol & Scale Requirement

| Physical Fabric | • 20–200+ Leaf Per Physical Fabric  
• Multi-Speed Interfaces on Same Leaf & Spine |
|-----------------|----------------------------------------------------------------------------------------|
| Traffic Throughput | • 1.5–5 Tbps  
• Linerate |
| Protocol | • BGPv4, BGPv6, Static route  
• BFD  
• IPv6 Multicast  
• ERSPAN with filter, SPAN on drop  
• Multi-Node PBR, PBR tracking, symmetric load-balancing, resilient hashing |
| Carrier Grade Scale | • 1500 Vlans  
• 50 VRFs  
• 1K Mac, ARP & ND per switch  
• 10–50K end hosts  
• 10–20K IPv4 & IPv6 LPM addresses  
• 3K–15K BGP & BFD across fabric |
Agenda

- Evolution of Telco Datacenter
- Customer Profiles & Technical Requirements

**How ACI met the requirements**
- Looking Ahead
- Conclusion
Application to Fabric Connectivity

vEPC deployment with separate Gi-LAN Fabric

- Separated ACI fabric for L4-L7 Services.
- Isolated change domain
- Gi-LAN fabric for wireless and wireline customers

Case Study #1
Telco cloud deployment for 4G

vEPC deployment with Cisco Ultra EPC and Cisco VIM

IP/MPLS

ASR9K

Spine N9364C

Leaf N93180YC-FX

Leaf N93180YC-FX

Leaf N93180YC-FX

Leaf N93180YC-FX

TCP Optimizer-1

TCP Optimizer-N

PGW, SGW, VoLTE, PCRF and other telco services running as virtual instances on Cisco VIM

vEPC service functions (data plane)

vEPC control functions (Control plane)
Distributed Edge DC with ACI Remote Leaf

- vEPC and Caching is deployed at Edge DC
- vEPC is connected to Remote Leaf
- APIC controller at Central site is managing Remote leaf deployed at Edge DC
Application to Fabric Connectivity

5G ready Central DC architecture

Spine
N9508
N97336-FX

Leaf
N93180YC-FX

Leaf
N93180YC-FX

Leaf
N93180YC-FX

Border Leaf
N9336C-FX2

Border Leaf
N9336C-FX2

vEPC
Cisco VIM

OSS/BSS
(Bare-Metal)

Gi-LAN
services

IMS
Cisco VIM

Big Data
(Bare-Metal)

Mobile Network Probe

NCS5508

NCS5508

IP/MPLS

Internet

Case-Study#3

200+ Leaf in a fabric
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
  • Automation
    • Services Connectivity
    • Operational Simplification
    • Faster Convergence
    • Distributed Edge DC
• Looking Ahead
• Conclusion
Fully Automated Provisioning of ACI Fabric

75% reduction in time spent bringing up network
• Fabric Provisioning
• Validation
• Inventory

Connect switches in Spine Leaf topology

Connect APICs to Leaf Pair
Power on APIC and switches

Input simple details like Fabric Subnet, APIC Out of Band Management IP & Login credential on APIC CIMC

Login to APIC and register switches

Fabric is up and running
Topology View
Automatic Policy deployment

Gateway- 100.1.1.1

Automatic deployment of Tenant, VRF, EPG, BD (Gateway IP) when End Point is detected

100.1.1.100
Vlan 100
Host A

100.1.1.101
Vlan 100
Host B

100.1.1.102
Vlan 100
Host C
Automation using APIs

Self developed Provisioning tool for pushing ACI policies as an alternative to UI based configuration

API Calls to provisioning interface policies, L3out, NTP and other one time polices

NSO/Self built too for provisioning
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
  • Automation
  • Services Connectivity
    • Operational Simplification
    • Faster Convergence
    • Distributed Edge DC
• Looking Ahead
• Conclusion
Evolved Packet Core (EPC) deployment
SAE GW Connectivity to ACI Leaf

Advertise subscriber pool to Border Leaf through BGP in fabric

Advertise subscriber pool static route, OSPF or BGP

Advertise subscriber pool outside of fabric

SAE GW (SGW & PGW)

BGP in Fabric

BGP RR

1.1.1.0/8 & 2000::/64
Subscriber Pool

IP/MPLS

Internet

ASR9K

1.1.1.0/8
2000::/64
vEPC characteristics

- vEPC VNF’s distributed across many Rack, Leafs
- Flexibility to deploy VNF’s based on capacity in Rack’s, Servers etc.

Virtual instances spread across multiple racks for vEPC solution

1.1.1.0/8 & 2000::/64
Subscriber Pool

IP/MPLS
Internet
vEPC traffic forwarding requirements

- ECMP within fabric towards leaf switches connected to VNF’s.
- ECMP from each leaf to all VNF’s hosted across multiple racks.
- Faster convergence using BFD

Each leaf having routing relations with VNF’s connected across multiple racks/leafs for ECMP. i.e Fabric wide ECMP with VNF’s

1.1.1.0/8 & 2000::/64 Subscriber Pool

Virtual machines spread across multiple racks for vEPC solution

IP/MPLS

Internet

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public
Cisco vEPC design with ACI

- vEPC functionality is divided across service functions (SF) and control functions (CF).
- CF advertises subscriber pool information through BGP with the common next-hop of user-plane VNFs (SF)
- All the use-plane VNFs are sharing a common IP
- ACI leaf switches are configured with static route or dynamic routing protocol for provide reachability to common IP of SF
- 64-way ECMP from ACI Leaf to SF across racks.
- BFD is used between SF and ACI Leaf for resiliency

SF= Service Function (Data Plane)
CF= Control Function (Control Plane)

CF-1
Subscriber pool
1.0.0.0/8
2000://64

Common IP for SFs
2.2.2.2
vEPC design with ACI

Design details

Static route or dynamic protocol with 64 way ECMP
2.2.2.2->10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4,….

SF= Service Function (Data Plane)
CF= Control Function (Control Plane)

BGP RR

BGP in Fabric

Subscriber pool
1.0.0.0/8->2.2.2.2 via BGP
2.2.2.2-> L1, L2, L3, L4

BGP with Control VM
Static route or dynamic routing protocol with BFD

Mobile Backhaul
Internet

Common IP for SFs 2.2.2.2
vEPC design with ACI

Packet walk

Equal traffic distribution among VNFs irrespective of the amount of VNFs connected to each Leaf switch

Subscriber pool 1.0.0.0/8 -> 2.2.2.2 via BGP
2.2.2.2 -> L1, L2, L3, L3

Static route or dynamic protocol with 64 way ECMP
2.2.2.2 -> 10.1.1.1, 10.1.1.2, 10.1.1.3, 10.1.1.4, ...

Mobile Backhaul

Internet

Common IP for SFs 2.2.2.2

SF= Service Function (Data Plane)
CF= Control Function (Control Plane)

BGP with Control VM
Static route or dynamic routing protocol with BFD
vEPC design with BGP from UPF with ACI

- 4G CUPS deployment for UPF
- Redundant and resilient design with BGP & BFD from each switch to SF
- Flexible Architecture, where UPF can be instantiated anywhere
- Each UPF is running BGP from UPF to all ACI leaf connected to vEPC cluster
- Increases overall BGP scale requirement on fabric
- 3000 BGP session across fabric is supported
- 15,000 BGP session across fabric (Q2CY20)
Load-balancing without GTP based load balancing

GTP Tunnel is formed between eNodeB and PGW. Source IP and destination IP of GTP tunnel would be PGW and eNodeB.

Since source IP and destination IP is fixed, traffic will always pick the same link between ACI Leaf and PGW. This may cause imbalance of traffic on one link of the PC member and packet drops.

SAE GW-1 (SGW & PGW)
1.1.1.1
1001::1

IP/MPLS

eNodeB
2.2.2.2
2002::2

1.1.1.0/8 & 2000::/64 Subscriber Pool
Load-balancing without GTP based load balancing

Since source IP and destination IP is fixed, traffic will always pick the same fabric links.

SAE GW-1 (SGW & PGW)
1.1.1.1
1001::1

1.1.1.0/8 & 2000::/64
Subscriber Pool

IP/MPLS

eNodeB
2.2.2.2
2002::2
GTP based load balancing

- Load-balancing based on GTP Tunnel id by looking at inner header on Spine
- Load-balancing based on GTP Tunnel id by looking at inner header on BorderLeaf
- Load-balancing based on GTP Tunnel id by looking at inner header on Server leaf towards Port-channel and Fabric

IP/MPLS

- eNodeB
- 2.2.2.2
- 2002::2

SAE GW-1 (SGW & PGW)
- 1.1.1.1
- 1001::1

1.1.1/8 & 2000::/64
Subscriber Pool
Gi-LAN services
Multi-Node Service chaining in Telco DC

SAE GW (SGW & PGW)

SAE GW (SGW & PGW)

Group of TCP Optimizers

CG-NAT

Deep Packet Inspection (DPI)

Gi-LAN

Subscriber Pool

Internet

ASR9K

© 2020 Cisco and/or its affiliates. All rights reserved. Cisco Public
Logical ACI Construct for service chaining

- **PGW EPG**: Classified based on subscriber Pool
  - 10.0.0.0/8
  - 2000::/8

- **Contract with Filters**: Protocol (ICMP, TCP, UDP), L4 ports

- **Service Graph Template**
  - Group of TCP Optimizers
  - CG-NAT
  - DPI

- **Internet EPG**: Classified based on Internet prefixes
Simplified Configuration

L4-L7 Service Graph Template - service-chaining

Consumer

EPG

TCP_OPT_

TCP_Opt

CG-NAT

DPI

Provider

EPG

TCP_OPT_Group Information
Route Redirect: true

CG-NAT Information
Route Redirect: true

DPI-Multi-Node Information
Route Redirect: true
PBR Bypass a node

SAE GW (SGW & PGW)

SAE GW (SGW & PGW)

Subscriber Pool

Group of TCP Optimizers

CG-NAT

Deep Packet Inspection (DPI)

Gi-LAN

Internet
Integration with TCP Optimizers
TCP Optimizer Integration with ACI

Automatic Load-balancing and Symmetry of traffic flow

- Symmetric PBR ensure return traffic chooses same TCP optimize
- Automatic load-balancing of traffic across different TCP optimizers based on forwarding table hash (Source IP, Destination IP, Source Port, Destination Port)
TCP Optimizer Integration with ACI

Tracking TCP Optimizer Liveliness

ICMP & TCP Tracking of Inside & Outside Interface

Automatic Load-Balancing to remaining TCP Optimizers after failure

Inside 1.1.1.1

Outside 2.2.2.1

N-TCP Optimizers in a group

Removes whole TCP Optimizer if either Inside or Outside interface goes down

TCP OPT 1

TCP OPT 2

TCP OPT 3

TCP OPT N

SAE GW (SGW & PGW)

SAE GW (SGW & PGW)

Subscriber Pool

Internet

ASR9K
TCP Optimizer Integration with ACI

Bypassing TCP Optimizers to avoid congestion

Traffic is directly send to internet when more than defined number of TCP Optimizers fails.

SAE GW (SGW & PGW)  SAE GW (SGW & PGW)

Subscriber Pool

Inside 1.1.1.1  Outside 2.2.2.1

TCP OPT 1  TCP OPT 2  TCP OPT 3  TCP OPT N

N-TCP Optimizers in a group

Internet

ASR9K
TCP Optimizer Integration with ACI

Non-Resilient Hashing

All flows get re-hashed on a PBR node failure, this can cause traffic drop for flows that lands on a PBR node that does not have a session information.
TCP Optimizer Integration with ACI

Resilient Hashing

Only the flows that were going through failed node gets re-hashed to one of the node
TCP Optimizer Integration with ACI
Resilient Hashing with standby node

Only the flows that were going through failed node gets re-hashed to standby node
TCP Optimizer Integration with ACI

Simplified expansion

New TCP optimizers can be added anywhere in fabric

SAE GW (SGW & PGW)

Subscriber Pool

TCP OPT 1
TCP OPT 2
TCP OPT 3
TCP OPT 4
TCP OPT 31

N-TCP Optimizers in a group

ASR9K

Internet
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
  • Automation
  • Services Connectivity
  • Operational Simplification
    • Faster Convergence
    • Distributed Edge DC
• Looking Ahead
• Conclusion
Operations Tools

- Topology Dashboard
- Troubleshooting Wizard
- End Point Tracker
- Faults
- Capacity Dashboard
- Link Statistics
- Health Score Card
- Traffic Map
- Upgrade/Downgrade
Troubleshooting Wizard

Visiblity & Troubleshooting

Faults
- Critical
- Major
- Minor
- Warning

Time Window
- From: latest 240 minutes
- To: now

Session Information
- Source: 10.1.1.100
- Destination: 10.2.1.100
- Type: Endpoint → Endpoint
Packet capture on APIC controller

- Easy to SPAN packets to APIC without dedicated server
- Packet policer to limit the number of packets sent to APIC
- Requires in band connectivity
- 10 PCAP files of 1MB each to limit the usage of APIC
SPAN based on filter

SPAN session on ACI that sends SPANs packet matching the 5-tuples filter

SPAN/ERSPAN on ACI based on 5-tuples filter
SPAN on drop

SPAN session on ACI that sends dropped packet to Analyzer
ERSSPAN from ACI fabric to Monitoring tools

- S1U traffic monitoring
- ERSPAN from each leaf to monitoring destination
- Dedicated rack for traffic monitoring tools
- Monitoring appliance could scale up to 10G traffic causing ERSPAN scale requirement on ACI fabric
- Increased ERSPAN scale from 4 bi-directional session to 16 bi-directional session per leaf
Simple Return Merchandise Authorization (RMA)

Decommission the old switch by removing the controller

Register the new switch by providing the same Node ID and Node Name.
Agenda

- Evolution of Telco Datacenter
- Customer Profiles & Technical Requirements
- How ACI met the requirements
  - Automation
  - Services Connectivity
  - Operational Simplification
- Faster Convergence
  - Distributed Edge DC
- Looking Ahead
- Conclusion
ACI Fabric Convergence

- **Controller Failure - No Loss**
- **Fabric Failure – 200 msec**
  Convergence happens within ASIC
- **Access Failure - within 200 msec**
  ACI Fabric to vPC connected host failure
- **External Connectivity Failure - within 200 msec**
  ACI Fabric to external connectivity failure

Internet
IP/MPLS
Active/Active Server
ASR9K
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
  • Automation
  • Services Connectivity
  • Operational Simplification
  • Faster Convergence
• Distributed Edge DC
• Looking Ahead
• Conclusion
Architecture for Distributed Datacenters

ACI Remote Physical Leaf

Remote Location contains Nexus 9300 connected to IP Network and fully managed by APIC cluster of Main DC

All local traffic is switched directly between endpoints, both virtual and bare metal
Architecture for Distributed Datacenters

Traffic forwarding between pair of Remote Leaf switches
Architecture for Distributed Datacenters

Direct traffic forwarding between pair of Remote Leaf switches
Distributed Edge DC with ACI Remote Leaf

- IP Network is only for Remote Leaf management purpose
- Each DC has local VRF that maps to common VRF on DC-Edge router. This ensures traffic always take IP MPLS/ WAN path instead of IPN (VXLAN) path
- Non-VXLAN (IP MPLS/ WAN) path between DCs allows customer to use different policies in transport
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
• Looking Ahead
• Conclusion
ACI Multi-Pod

Single Telco DC Campus with multiple server halls

- Managed by a single APIC Cluster
- Single Management and Policy Domain
- End-to-end policy enforcement
- Control plane fault isolation
ACI Multi-Pod with transport network slicing

- Multi-Pod with each Pod in different location, and with a requirement of network slicing in the transport
- ACI Multi-Pod solution is only used for management of multiple Pods in this scenario
- Traffic between Pod is forwarded through WAN and not through IPN
Management of Multiple Sites with ACI Multi-Site Solution

- Separate ACI Fabrics with independent APIC clusters
- ACI Multi-Site pushes cross-fabric configuration to multiple APIC clusters providing scoping of all configuration changes
- End-to-end policy definition and enforcement
ACI Multisite with transport network slicing

- Traffic between ACI site is forwarded through WAN
- No ISN and No need for VXLAN stretch between ACI Sites
- Multi-site Orchestrator is pushing policy configuration across sites along with day-2 operations
Automation of Transport and DC using NSO

- NSO DC core function pack (DC-CFP) for DC to push policies on ACI
- NSO transport software defined network core function pack (T-SDN CFP) to push policies to transport devices
- Cross-domain integration between transport and DC to automate configuration such as VRFs, QOS/SR label mapping, route-target configs etc.
Nexus Data broker for monitoring

- Advance ingress packet filtering and modification
- Load-balancing in egress

Filter traffic based on 5-tuples (source IP, destination IP, protocol, source port, destination port) filter in ACI

SPAN/ERSPAN from ACI towards Nexus Data Broker (NDB) switches

Analytics and Monitoring Tools
ACI-SR Handoff at Border Leaf

- ACI to SR Interworking at ACI border leaf
- Remote Leaf can be border leaf and support same design
- ACI to SR MPLS Interworking with N9300-FX2
- APIC/MSC to WAN controller integration
Cisco Network Assurance and Insights

Assure intent
“Ensure the business needs are consistently maintained”

Guarantee Reliability
“Solve problems before they impact business”

Troubleshoot intelligently
“Highlight the needle in the haystack”
Network insights - Enable proactive action
Increase availability, performance, and simplify operations

Sources Of Telemetry Data
- Config file
- Syslog
- Tech-support
- RIB
- FIB
- Accounting logs
- Debug logs
- Streaming telemetry
- Environmental
- Event history
- Cores
- Consistency checkers

Ingest And Process
- Metadata extraction
- Correlate against dBase
- Complex correlation

Derive Insights
- Anomaly Flows
- Predictive failure
- Root cause

Suggest Action
Network Insights – Use Cases

- Identify, Locate, Root cause, Remediate
- Error detection, latency, Packet drops
- Control plane issues
- Automated alerts
- Visibility
- Capacity alerts
- Environmental alerts
- Correlation
- Statistical models
- Upgrade impact
- Advisories
- Mitigate
- Prevent outages
- Hardening checks
- Software Hardware recommendations
- PSIRT notices
- EoS/EoL notices
- TAC assist
- Topology checker
Agenda

• Evolution of Telco Datacenter
• Customer Profiles & Technical Requirements
• How ACI met the requirements
• Looking Ahead
• Conclusion
Conclusion

- Massive Scale
- Time to Market
- Operations
- Service chaining

Looking Ahead
- Automation
- Distributed DC
- Scale
- Simplified Operation
- Integration with Tools

Consistent Policy & Management across Geography
Network Insight & Assurance
Please complete your session survey after each session. Your feedback is very important.

Complete a minimum of 4 session surveys and the Overall Conference survey (starting on Thursday) to receive your Cisco Live t-shirt.

All surveys can be taken in the Cisco Events Mobile App or by logging in to the Content Catalog on [ciscolive.com/emea](http://ciscolive.com/emea).

Cisco Live sessions will be available for viewing on demand after the event at [ciscolive.com](http://ciscolive.com).
Cisco Webex Teams

Questions?
Use Cisco Webex Teams to chat with the speaker after the session

How
1. Find this session in the Cisco Events Mobile App
2. Click “Join the Discussion”
3. Install Webex Teams or go directly to the team space
4. Enter messages/questions in the team space
Continue your education

- Demos in the Cisco Showcase
- Walk-In Labs
- Meet the Engineer 1:1 meetings
- Related sessions
Thank you
You make possible