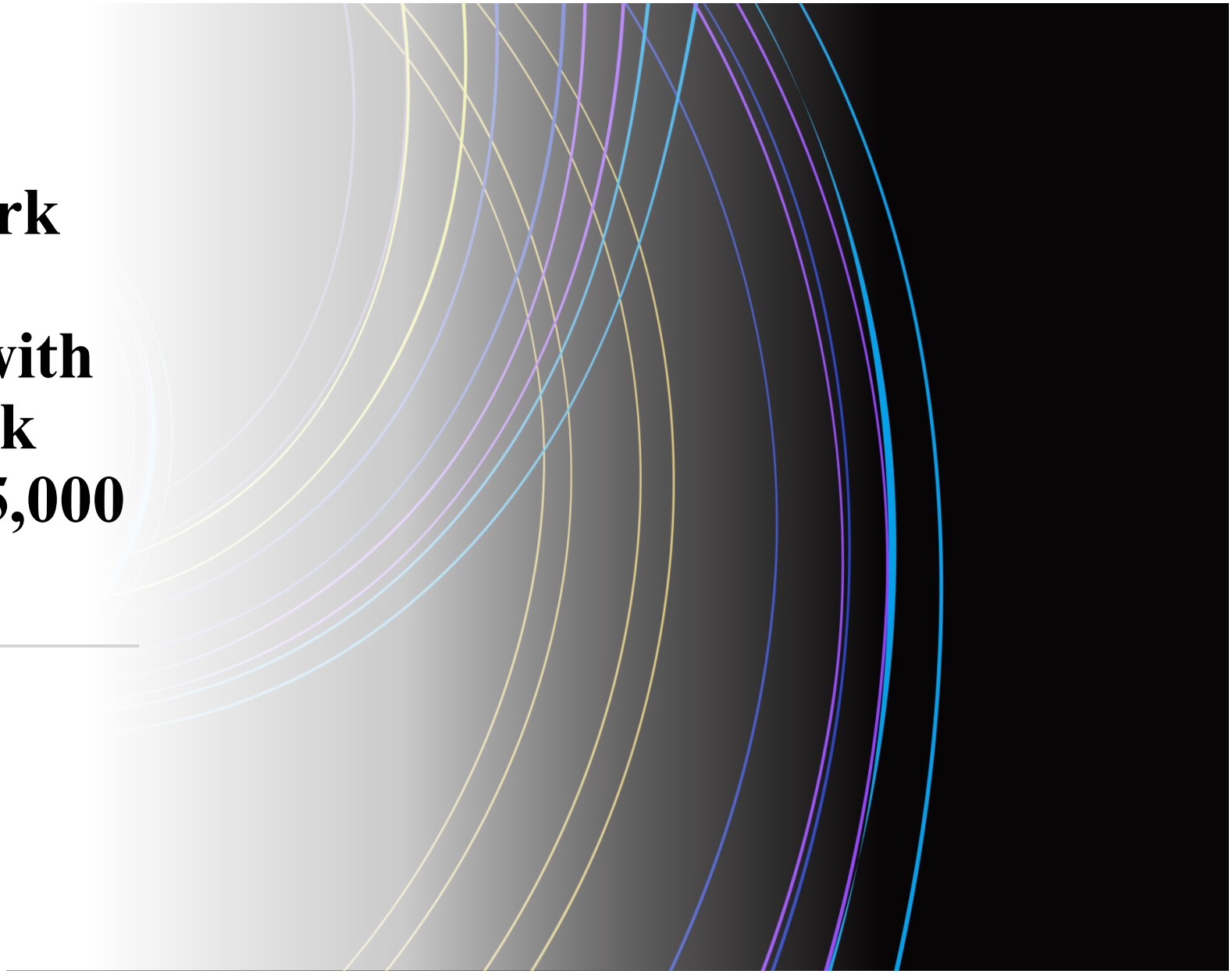
# Scaling Network for about 50 million users with 10,330 Network Devices and 55,000 km of Fiber

Sumon Ahmed Sabir
CTO, Fiber@home Limited
*sumon@fiberathome.net*



# Bangladesh

- Population 170 million
- Area  148K SqKM
- Mobile Internet user 114 million
- Broadband user 10.1 million


# Bangladesh : Progress in last 12 Years

| | 2010 | 2022 |
|---|---|---|
| Internet Users | 3.6 Million | 124.42 Million |
| Mobile Internet Users | 3.6 Million | 112.55 Million |
| Broadband Users | 12,000 | 11.87 Million |
| Bandwidth Consumption | 7.5 Gbps | 4,450 Gbps |
| Bandwitdth Price per mbps | BDT 27,000 | BDT 285 |
| Nationwide Optical Fiber Network | 15,000 Km | 1,58,000 Km |

# Fiber@Home Network at a Glance

| DWDM Devices | Count |
|---|---|
| Huawei | 518 |
| Infinera | 41 |
| Nokia | 51 |

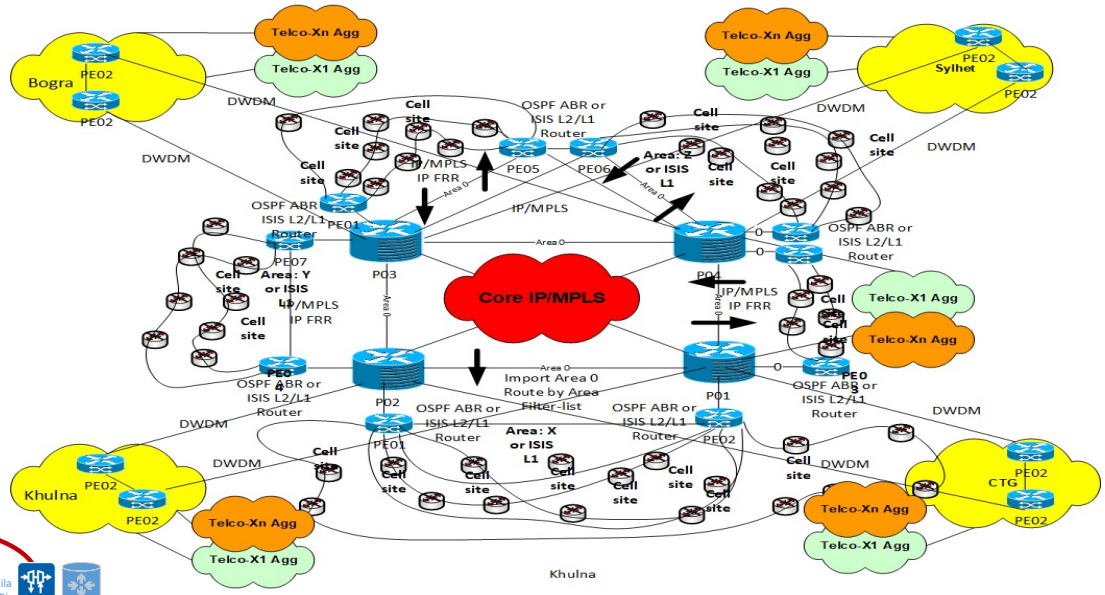
| IP Network Devices | Count |
|---|---|
| Cisco Routers | 4868 |
| Huawei Routers | 3149 |
| Huawei Switchs | 218 |
| Cisco Switchs | 62 |
| OLTs(Huawie/Alcatel/LS Cable) | 29 |
| Other vendors | 4 |
| Total | 8330 |

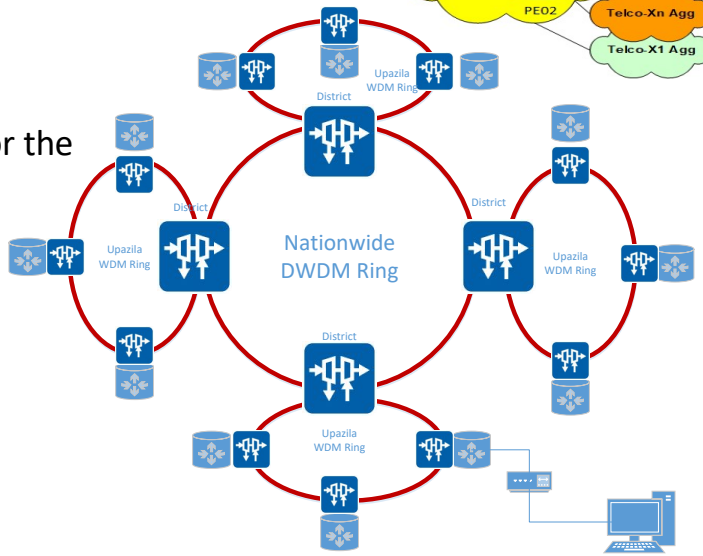

54,500 Km of Fiber Network Carriying 40% of nationwide traffic

# Fiber@Home Services

**Services:**

- L3VPN, L2VPN
- E1/ SDH over MPLS
- Dark Core
- Ethernet over DWDM
- Internet Transit
- Co-Location Service/Data Center
- Public Cloud is coming up




- Backhaul for ISPs and Telecos
- L2/L3 VPN services across the country for the Enterprises and Government offices
- IP RAN Backhaul for 2G/3G/4G/5G
- DC-DR Connectivity
- PON to home in limited area

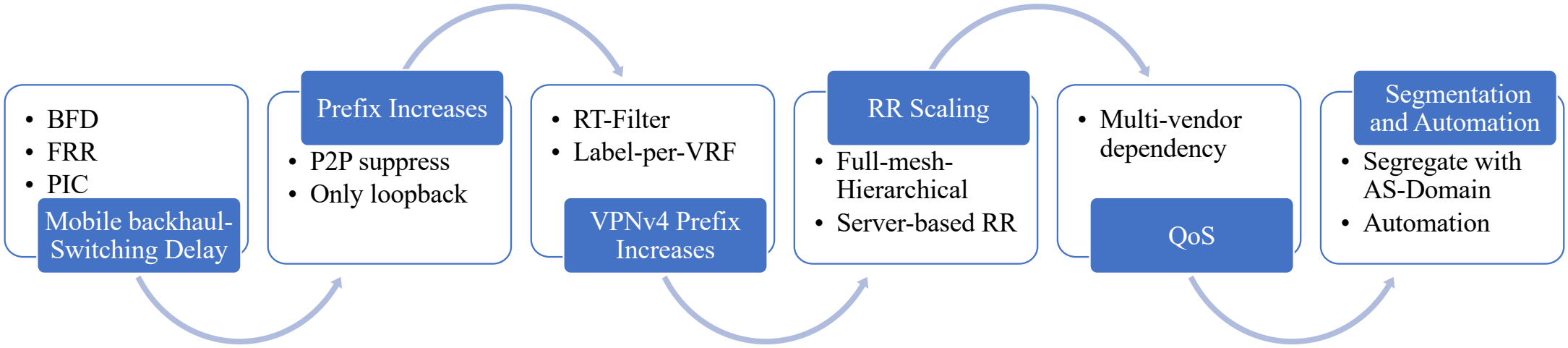
**Customers:**

- Telco- GrameenPhone, Robi, Banglalink, Teletalk
- ISP, IIG, IGW, IPTSP, etc.
- Govt Organizations

# IP Network is simple!

- Configure Interfaces, Run IGP, iBGP, eBGP with your clients and upstreams, run MPLS in transport network, Configure l2VPN, L3VPN, EVPN, You are ready to go.
- But as network grows Complexity starts!

# Network is Getting Complex with the Growth



# Need for Fast Convergence Emerges

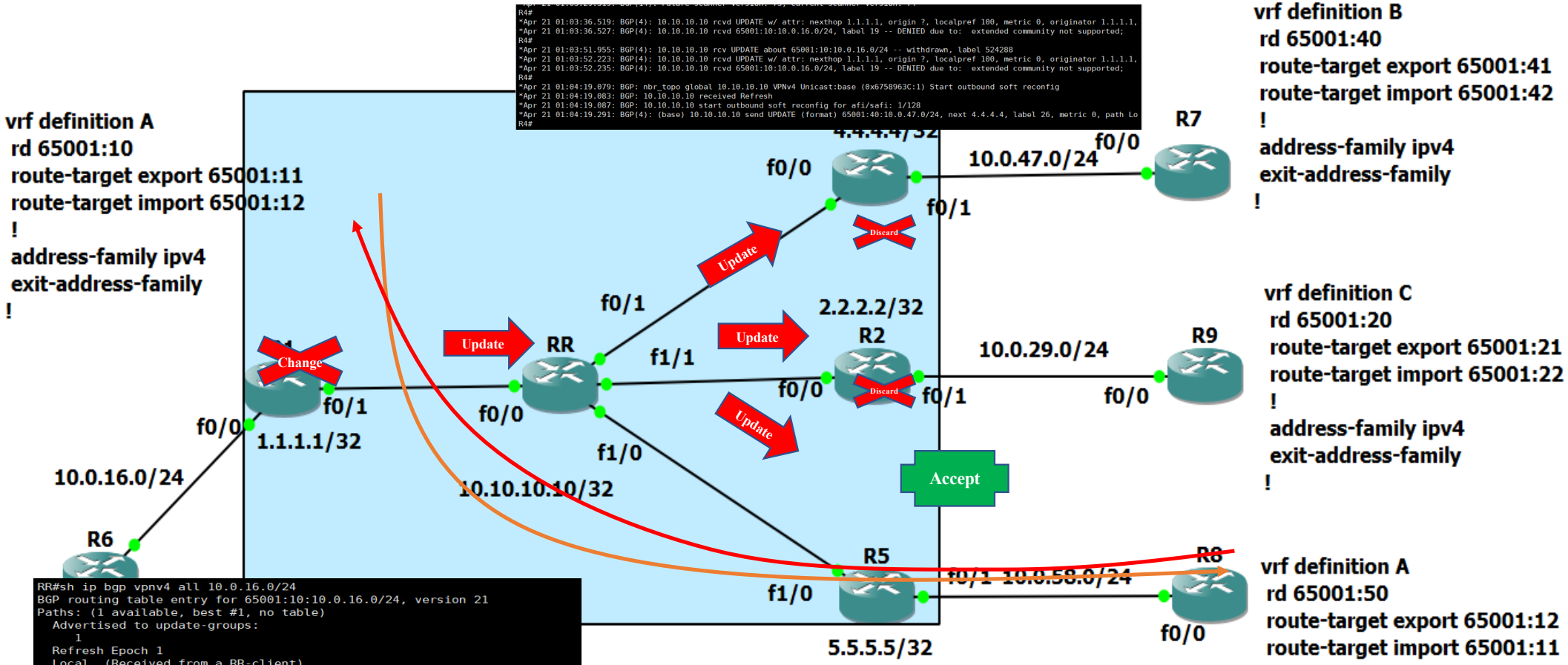
- In case of link failure need to avoid call drops/packet drops
  - Leads to IGP/BGP parameter tuning
  - Added BFD for fast identification of link failure
  - Added LFA-FRR for loop avoidence for transient interval in Access Ring Topology
  - Added BGP PIC to insert backup path for L3VPN Prefixes

- After all these we have achieved almost a Zero Packet drop network in case of link failure

# Network Grows on : 2000+ Routers

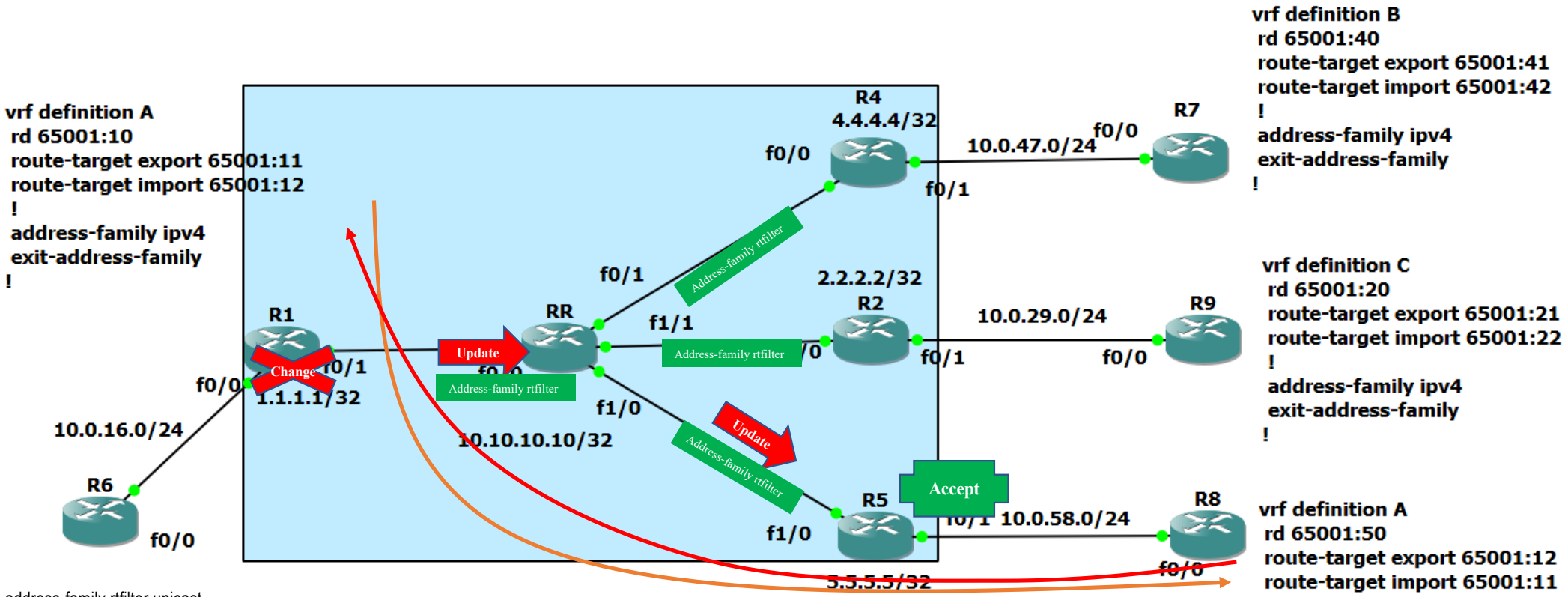
- Number of IGP prefixes become too high for the small cell site routers
  - Lead to filtering P2P link prefixes in Inter area IGP
  - Only Loopback IP’s are allowed to pass through

# Access Routers become slow again with the Growth




Any change in network affects the Cell site routers

# Use of Router-Target Filter




```
address-family rtfilter unicast
 neighbor 10.255.255.107 activate
 neighbor 10.255.255.107 send-community both
 neighbor 10.255.255.126 activate
 neighbor 10.255.255.126 send-community both
exit-address-family
```

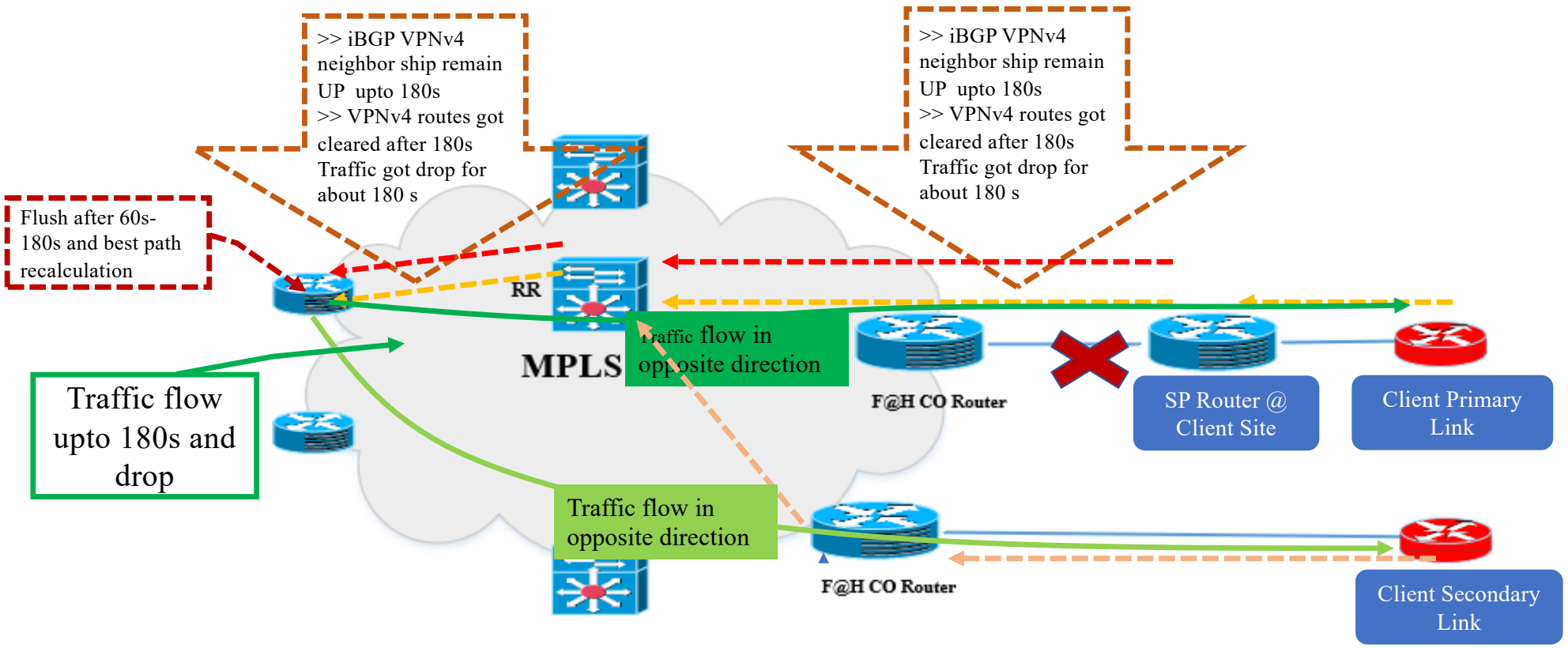
**Issues:**

- Change in any VRF are propagated to all PEs where there is no RT value
- This is a wastage of bandwidth, process and affect faster convergence

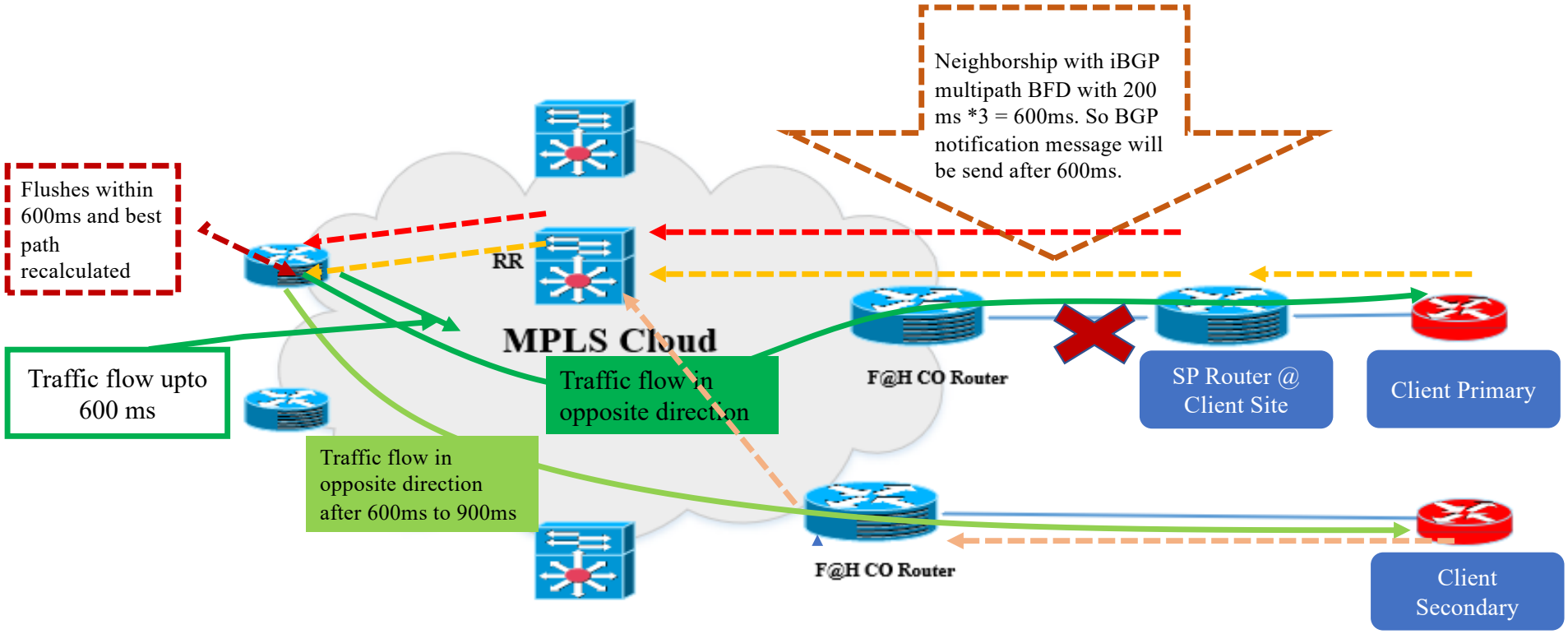
**Solution:**

- RT-Filter will ensure changes of only intended VRFs will be propagated

# Switching Delay 60s – 180s in Some Specific Topology

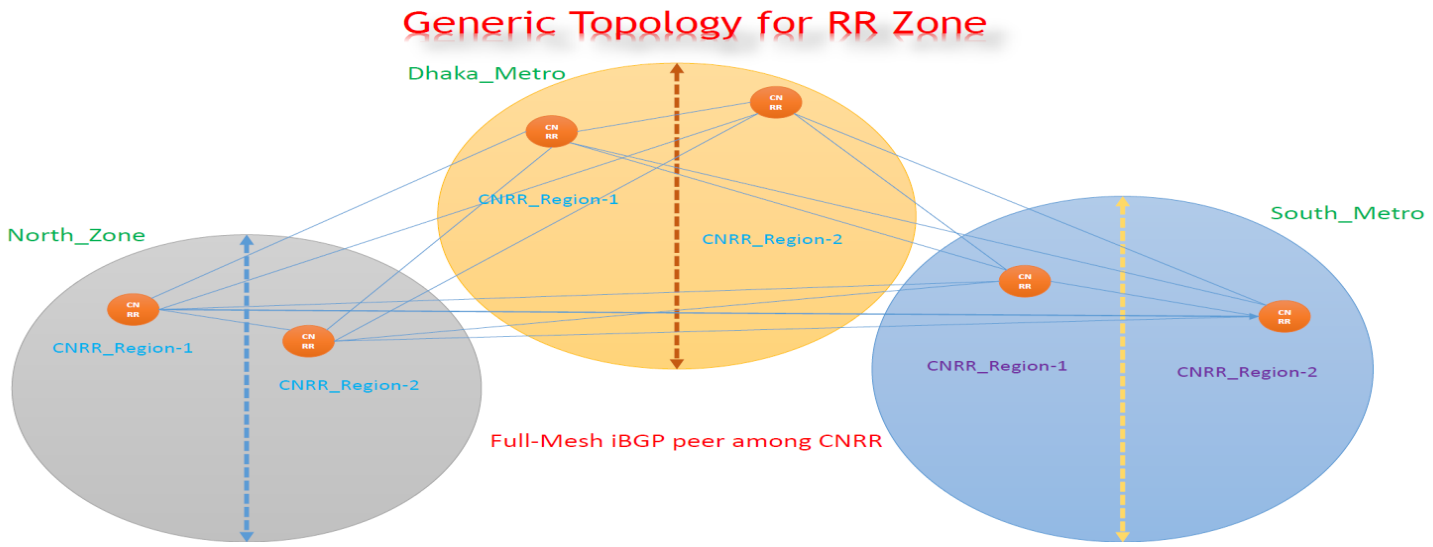


# Multi-hop BFD in BGP for Faster Convergence



# Challenges with RR Design and Solutions

Mesh Design >> Common Design




**Benefits:**

- Simplified design, configuration and maintenance
- Easy to implement

**Applicable for:**

- Small network
- Neighbor and prefix count is not large

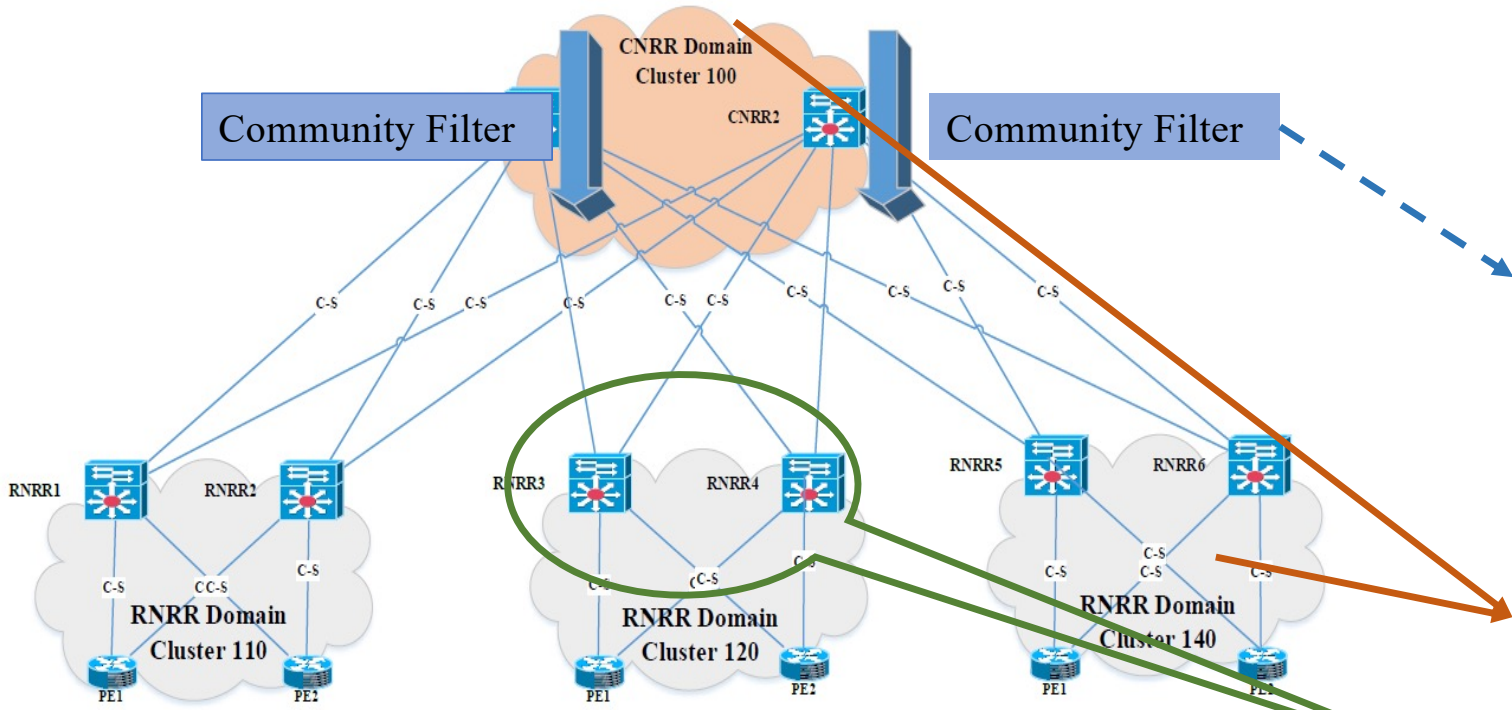
**Problem:**

- With the increase of devices/ prefixes - problem may start.
- Prone to “Routing Information Loop”/ “Switching Delay” due probable loss of synchronization and huge overhead processing delay when neighbor and prefix count increases.

**Solution:**

- Hierarchical design with optimization techniques

# Hierarchical RR Design : RTC, Cluster ID, Community




Mesh to Hierarchical Design

Community Filter >> to reduce VPNv4 signaling at CNRR-RNRR direction

Cluster-ID >> to prevent any routing information loop

RT-Filter >> reduce VPNv4 signaling

# Router based RR vs Server based RR Solution

**RR Upgrade:** Existing RRs are of old model Routers with low memory and processing power and in need for a replacement.

**Wastage of Forwarding Plane Capacity:** For RR, we mainly require control panel capacity. But routers are developed control plane + forwarding plane capability where later is wasted.

**Cost:** Need a huge budget for router-based RR deployment, upgrade and maintenance compared to server-based RR.

**Current Industry Practice:** Many service providers have already moved to server-based RR

| | USD |
|---|---|
| Router RR (BGP neighbor- 5000, Prefix- 25M) | 26500 |
| Server-Based RR of same capability | 5000 |
| Price Comparison | 530% |

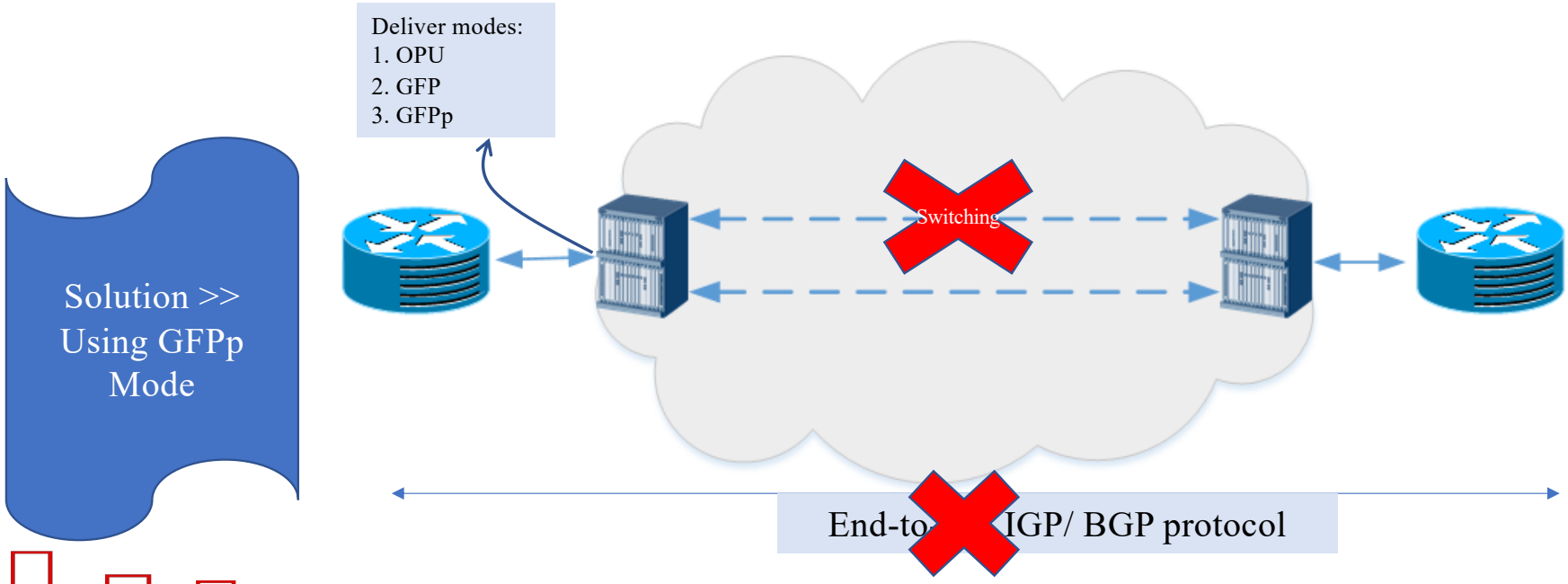
# Challenges with IP over DWDM Network

**Design:**

- IP over DWDM as transport network ensures high resiliency and redundancy.
- DWDM has different client deliver modes- OPU, GFP, GFPp
- Faster convergence and BW utilization: OPU > GFP > GFPp.

**Problem:**

- During switching in DWDM network, end-to-end protocol in IP domain gets down even after setting carrier delay.



# QoS Deployment in Multi-Vendor Environment

**Why QoS:**

- To protect sensitive and priority traffic during congestion time.

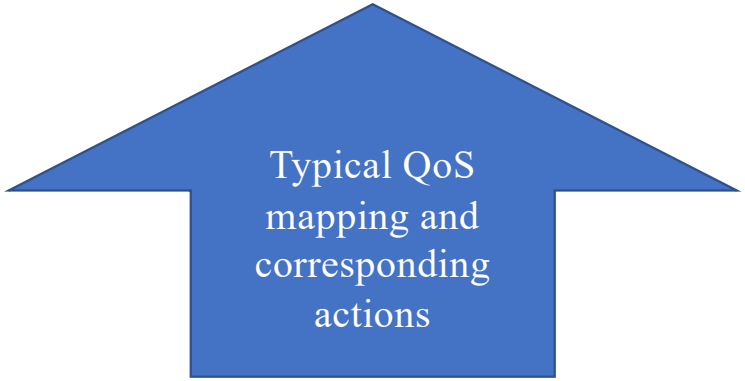
**Problems:**

- In a multi-vendor MPLS network, DSCP mapping, marking and actions are different and we don't get proper functionalities with the vendor-defined default behavior.

| Input DSCP | Input DSCP Decimal Value | MPLS EXP | Cisco Retained DSCP after MPLS mapping/ de-mapping | Fixing By | Huawei Retained DSCP after MPLS mapping/ de-mapping | Fixing By |
|---|---|---|---|---|---|---|
| CS0 | 0-7 | 0 | CS0 | changing "mpls uniform mode" to "pipe mode" | CS0 | changing "mpls uniform mode" to "pipe mode" |
| CS1,AF11,AF12,AF13 | 8-15 | 1 | CS1 | | AF11 | |
| CS2,AF21,AF22,AF23 | 16-23 | 2 | CS2 | | AF21 | |
| CS3,AF31,AF32,AF33 | 24-31 | 3 | CS3 | | AF31 | |
| CS4,AF41,AF42,AF43 | 32-39 | 4 | CS4 | | AF41 | |
| CS5,EF | 40-47 | 5 | CS5 | | EF | |
| CS6 | 48-55 | 6 | CS6 | | CS6 | |
| CS7 | 56-63 | 7 | CS7 | | CS7 | |

**Solution:**

- Designing QoS with lower-level control, precise and granular end-to-end mapping, marking and actions remains so that these actions remains same in different platforms of different vendors.
- Customized buffer assignment depending on traffic class.

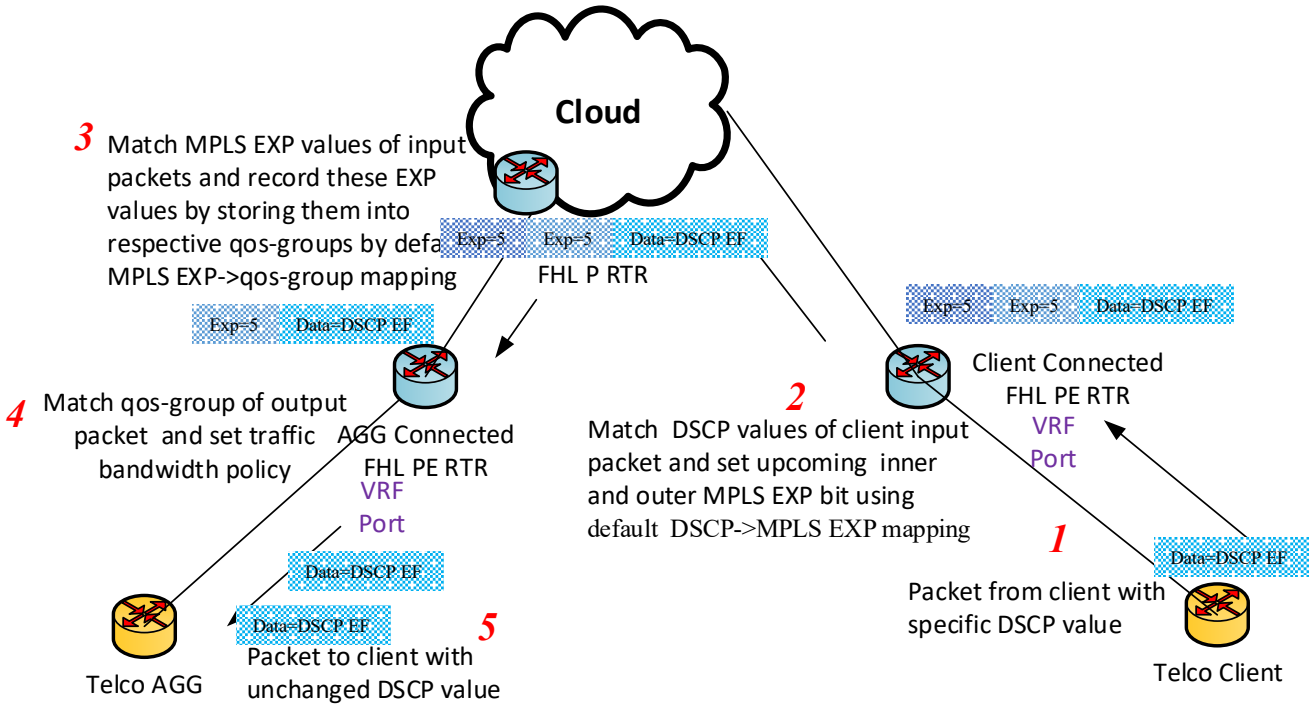


Location A (e.g. Branch Office)
→ Location B (e.g. Main Office)
**Default Mapping**

# Default Mapping

(considering DSCP change issue resolved by pipe mode- label per vrf)

Challenges with QoS Deployment and Functionalities in Multi-Vendor Environment



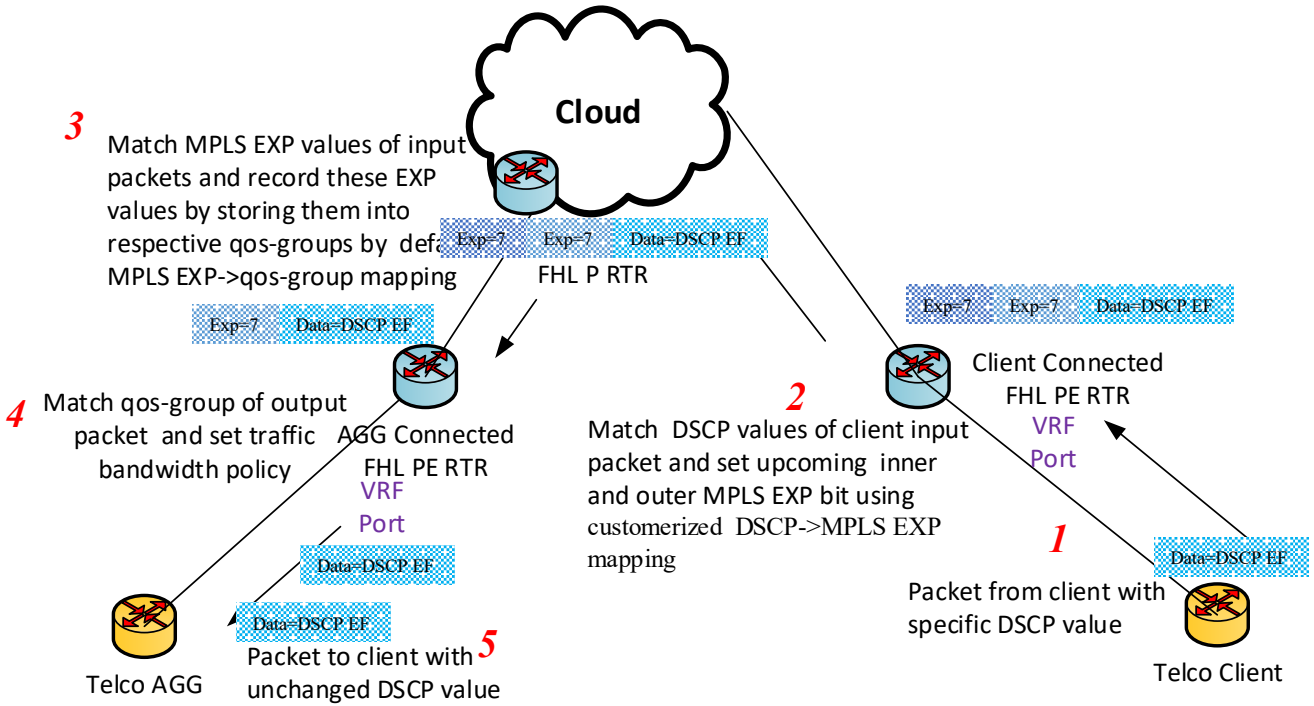

- EF is being mapped to exp5 in mpls but target is to set to exp7 due to business requirement.

Location A (e.g. Branch Office)
→ Location B (e.g. Main Office)
**Customized Mapping**

# Customized Mapping

Fiber @ Home

**Challenges with QoS Deployment and Functionalities in Multi-Vendor Environment**

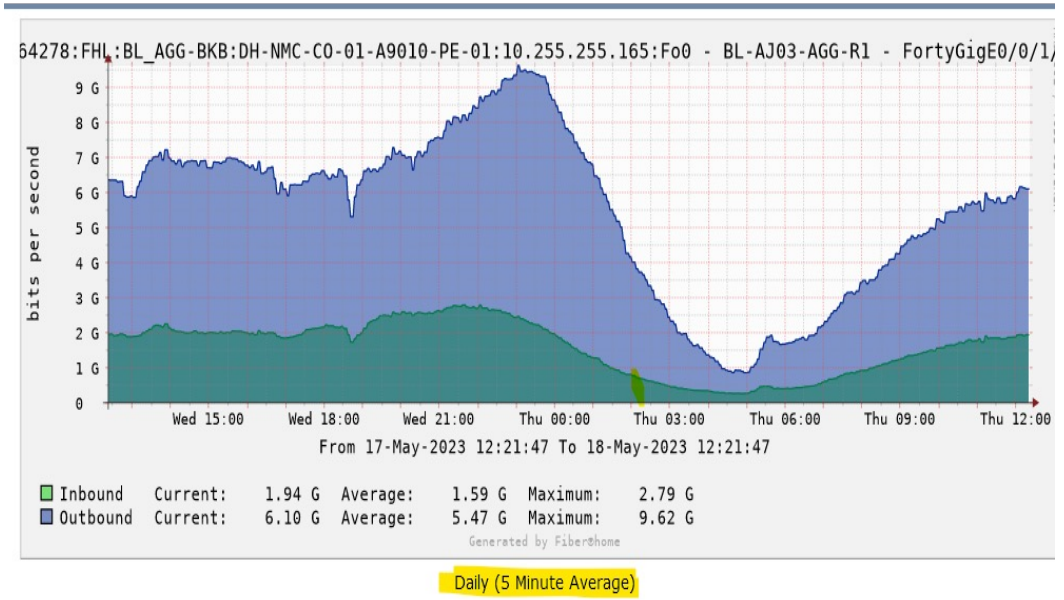



- Done customized mapping, marking to achieve expected mapping EF to mpls exp7

# Challenges with Measuring Actual Interface Utilization

**Problems:**

- Cacti/ MRTG etc. works on SNMP-based polling with a preferred polling frequency 5 min. Can be reduced to 1 min but still it's not good enough and needs huge resource.
- Proper utilization is not being captured as utilization is of average nature.
- Peaks and bursts are not being identified.
- Has commercial impact
- Impacts network upgradation planning
- Impact troubleshooting and buffer setting




Daily (5 Minute Average)

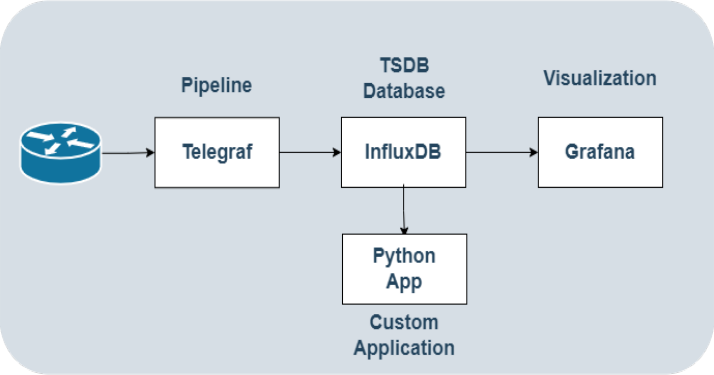
**Expectations:**

- Second or sub-second level utilization to see the actual utilization

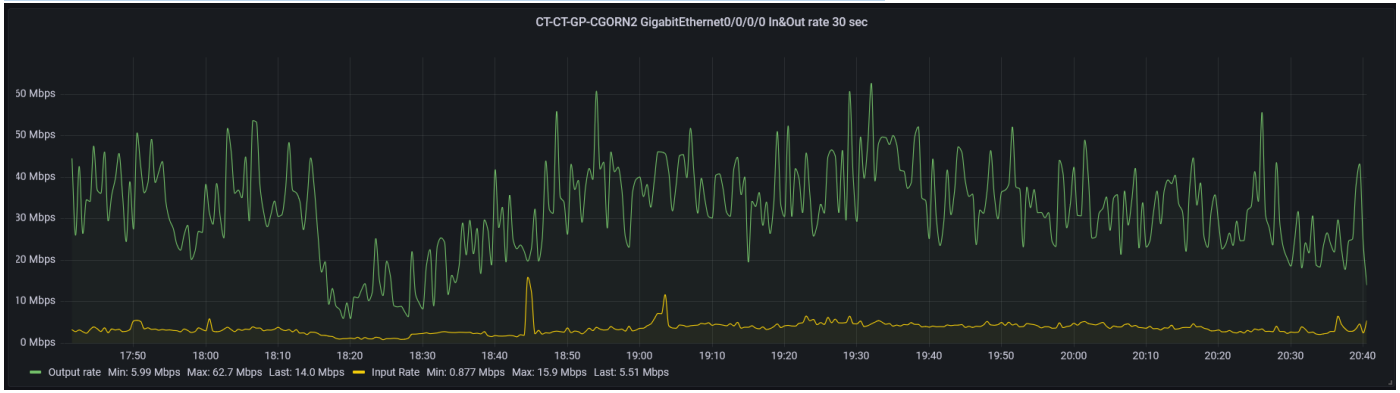
**Solution:**

Telemetry-based TIG stack >> next slide

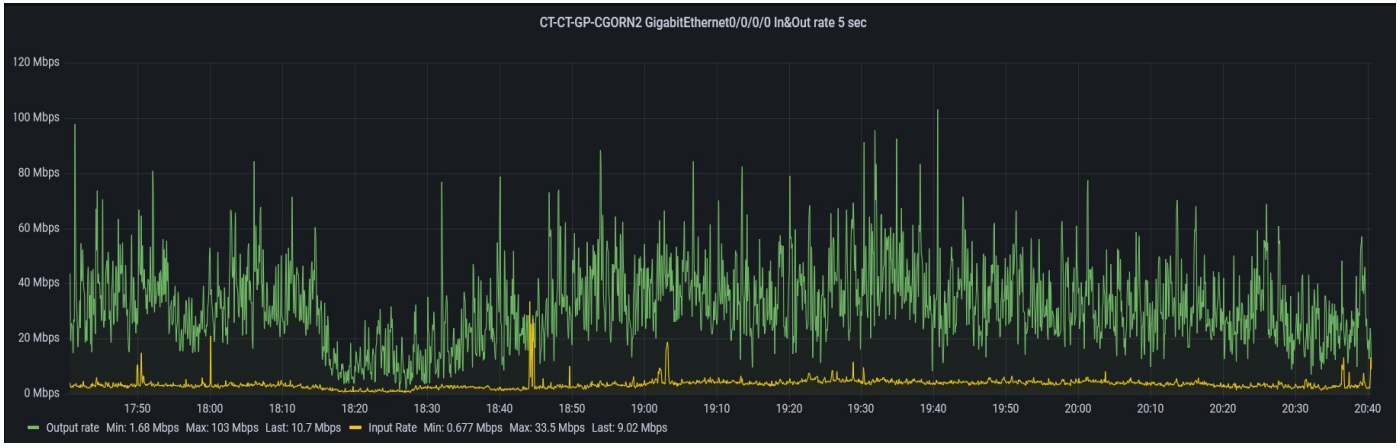
# Measuring Actual Interface Utilization




30 sec Inerval >> max utilization 15 mbps




5 sec Interval >> max utilization 33 mbps




**Benefits:**

- Helps to find the actual utilization
- Network upgrad
- Troubleshooting and buffer assignment
- Business opportunity

# High-Loss Alarm for Link Budget Threshold : TIG Solution

**Problem:**

- NMS generate alarm based on system-defined threshold breaching.
- Generating alarm based on link-budget threshold is highly challenging with NMS.

**Solution:**

- Telemetry-based TIG stack, Django or other web application

Threshold Info in Port Description → Optical power collection by TIG stack → Backend script and web-app to generate and view alarm

- A Python script is there to retrieve real-time optics data from InfluxDB.
- Generates alarm by verifying the threshold value of each link
- Django web application to create TT automatically and for NOC visibility.

Optical Power Alarm Management(High Loss)

| Total links in monitoring: | Total device in monitoring: | Total Alarm(s): | Cleared Alarm(s): |
|---|---|---|---|
| 5758 | 1112 | 60 | 1991 |

Choose link type: All  Choose Operator: All

Excel  Search:

| Serial | Alarm ID | Priority | Alarm Status | ACK Status | Alarm Time | IP | Interface | Hostname | TX | RX | THG |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 60 | 202304180097 | 4 | RUNNING | | April 18, 2023, 8:43 p.m. | 10.253.231.231 | GigabitEthernet0/0/0/5 | DH-GU-RB-DHGULP6 | -30.45 | -32.21 | -19.0 |
| 59 | 202304180093 | 4 | RUNNING | | April 18, 2023, 7:32 p.m. | 10.253.105.90 | GigabitEthernet0/0/0/17 | RB-CTG-CHOWKBAZAR_WIC | -6.66 | -28.23 | -26.0 |
| 58 | 202304180074 | 3 | RUNNING | | April 18, 2023, 4:19 p.m. | 10.253.199.35 | TenGigE0/0/0/22 | SA-SHARIATPUR-CL-01-N540X2C-PE-01 | 1.49 | -23.56 | -23.0 |
| 57 | 202304180069 | 4 | RUNNING | | April 18, 2023, 3:58 p.m. | 10.253.148.39 | GigabitEthernet0/0/0/6 | NW-SH-RB-NWSBG04 | -6.44 | -33.01 | -11.0 |
| 56 | 202304180046 | 2 | RUNNING | | April 18, 2023, 11:57 a.m. | 10.255.255.189 | FortyGigE0/0/1/0 | MU-SREENAGAR-CL-03-N5402C-PE-01 | 3.11 | -18.79 | -18.0 |
| 55 | 202304180026 | 4 | RUNNING | | April 18, 2023, 7:13 a.m. | 10.255.255.111 | GigabitEthernet0/0/0/7 | RS-RAJ-CL-02-N5402C-PE-01 | -7.44 | -18.82 | -18.0 |
| 54 | 202304180023 | 3 | RUNNING | | April 18, 2023, 6:56 a.m. | 10.253.165.236 | TenGigE0/0/0/11 | ST-SATKHIRA-CL-01-N540X2C-PE-02 | 2.65 | -18.26 | -18.0 |

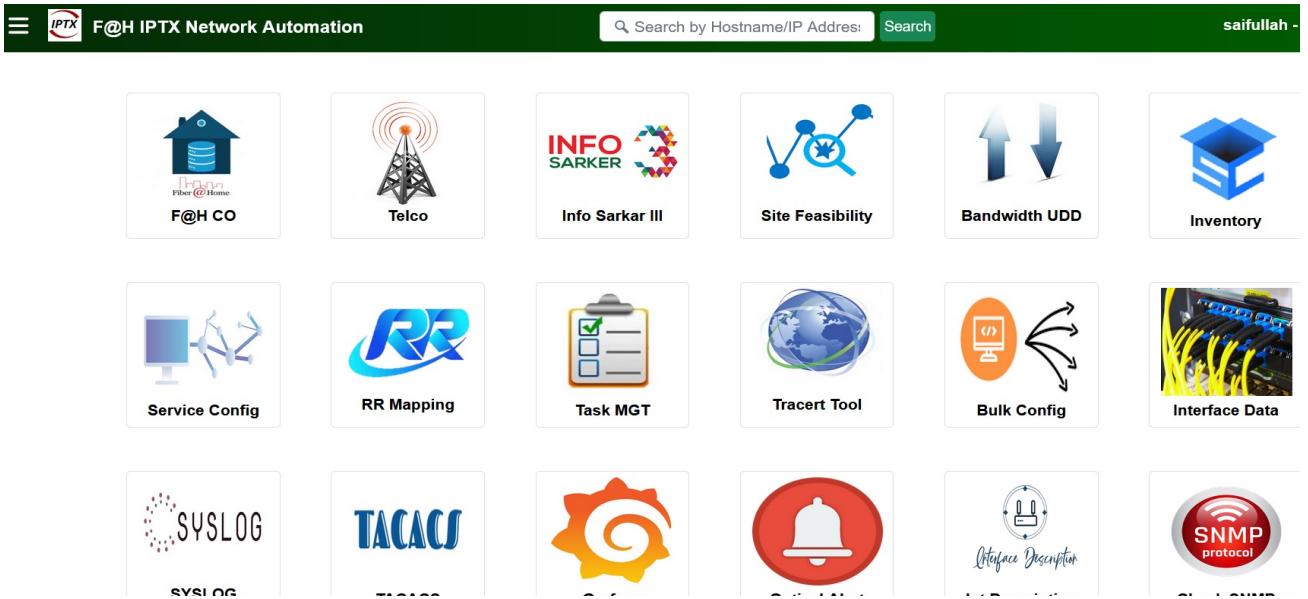
# Automation : Service Configuration with Visibility : Python, Ansible

**Challenges:**

- Getting customized network information, visibility, inventory management etc.
- Manual feasibility analysis
- Configuration during big project roll-out
- Huge manhour is required for (1) manual configuration (2) network audit, (3) health check etc.
- Manual works are also error-prone.
- Vendor solution is costly.
- AoB

**Solution:**

- Network programmability with Python scripting- Telnetlib, paramiko, netmiko, napalm, netconf, restconf etc.
- Web-based customized tool e.g. DJango for task management




**BKB Link Utilization Summary**

| Link Type | Total Link | (0 to 39)% | (40 to 59)% | (60 to 79)% | 80% Above | DWDM_Link | Fiber_Link |
|---|---|---|---|---|---|---|---|
| EXPRESS BKB | 54 | 15 | 21 | 13 | 5 | 54 | 0 |
| L3 CORE BKB | 603 | 529 | 45 | 22 | 7 | 67 | 536 |
| L2 CORE BKB | 17 | 17 | 0 | 0 | 0 | 2 | 15 |
| D2D BKB Link | 590 | 520 | 45 | 17 | 8 | 0 | 590 |
| OLT Uplink | 58 | 55 | 2 | 0 | 1 | 0 | 58 |
| NTTN IS3 NNI | 53 | 32 | 9 | 10 | 2 | 0 | 53 |

# Network Segmentation : Inter-AS Routing

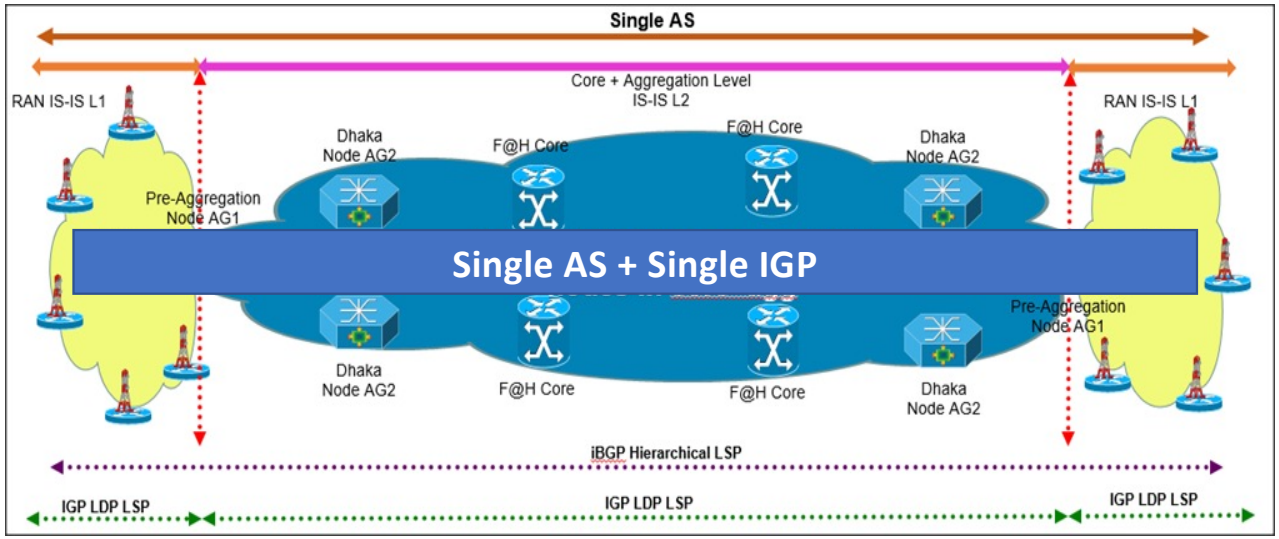
## Network Statement

- Single As with Multi Area
- BGP LU Network is running

## Problem Statement

- Lack of control on traffic flow
- Sub-optimal routing in some multi-exit ABR Router
- Lots of challenges to adopt new technology
- Unmanaged network growth

## Current Architecture



# Network Segmentation : Inter-AS Routing

## Network Segmentation

### Network Segregate

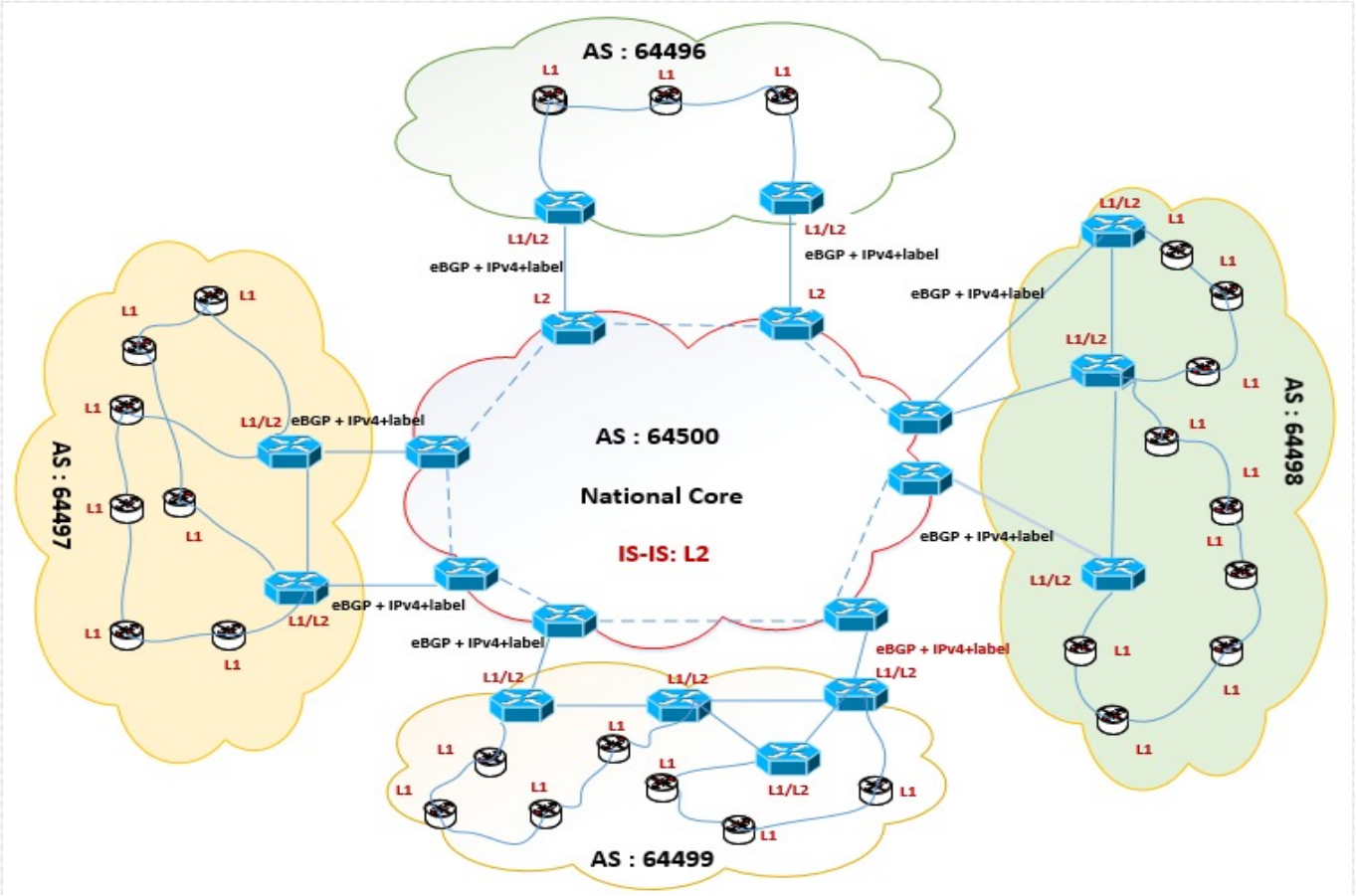
- Single Network separate by multiple AS
- National P will have ISIS: L2 Signal
- National P will connect with Zonal ASBR

### Zonal PE

- Zonal PE will have L1/L2 signal
- Zonal PE could have ring between two or more CO within AS.
- Zonal PE to P will communicate through eBGP + IPv4+label

### CSR

- All CSR will have L1 Signal
- Non protocol link will connect through Switch



# SR-MPLS vs SRv6

- Some vendors are well-prepared for SR-MPLS, very few are for SRv6. What could be be migration strategy in a multi-vendor environment?
- What are the deployment and operational challenges may appear?

# Network is Still Growing

- Expecting another 2000+ router addition in coming year.
- Probably will bring newer challenges!

# Thank you!