Accelerated Open Source vBRAS Solution Based on OpenBRAS and VPP/DPDK

Hongjun Ni & Liang Ou, Yujia Luo
Intel China Telecom

Acknowledgement:
Heqing Zhu, John DiGiglio, Beilei Xing @Intel
Neale Ranns @Cisco

.io Foundation

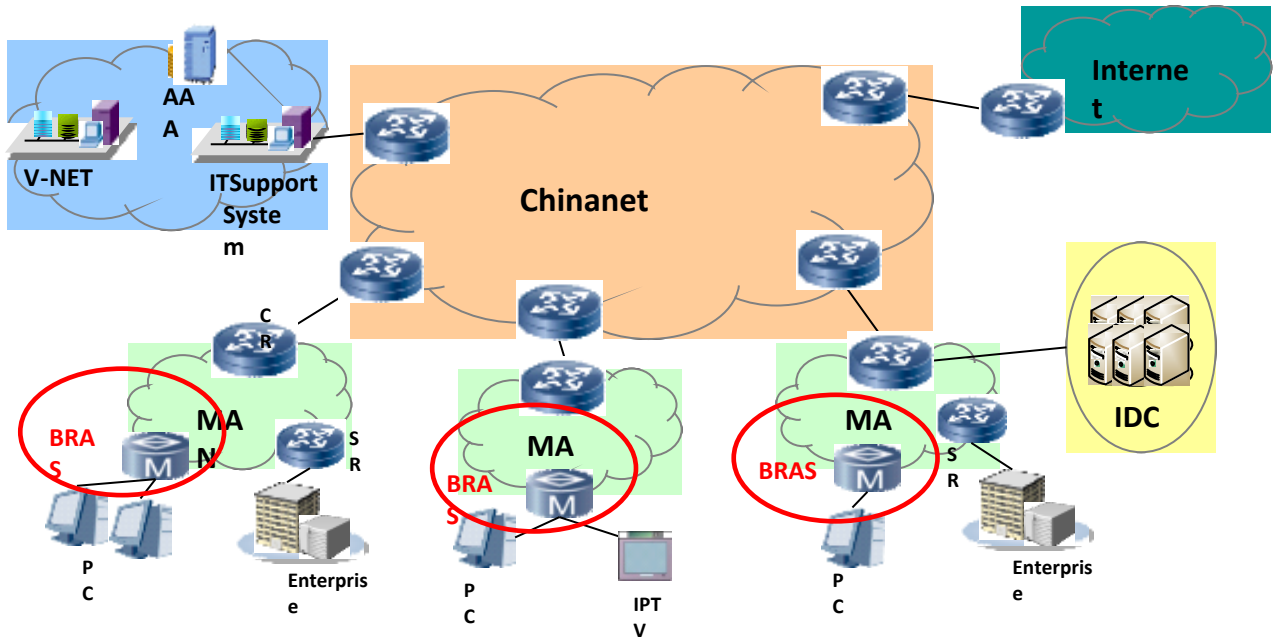
Agenda

- BRAS in China Telecom
- OpenBRAS Architecture
- vBRAS Deployment
- Data Plane implementation
- Distribute traffic using NIC
- Key Takeaway

BRAS in China Telecom

- **What is BRAS ? B**roadband **R**emote **A**ccess **S**erver
  - ✓ authenticate, authorize and route broadband subscriber data traffic according to customized policies

- **Where does BRAS locate ?**



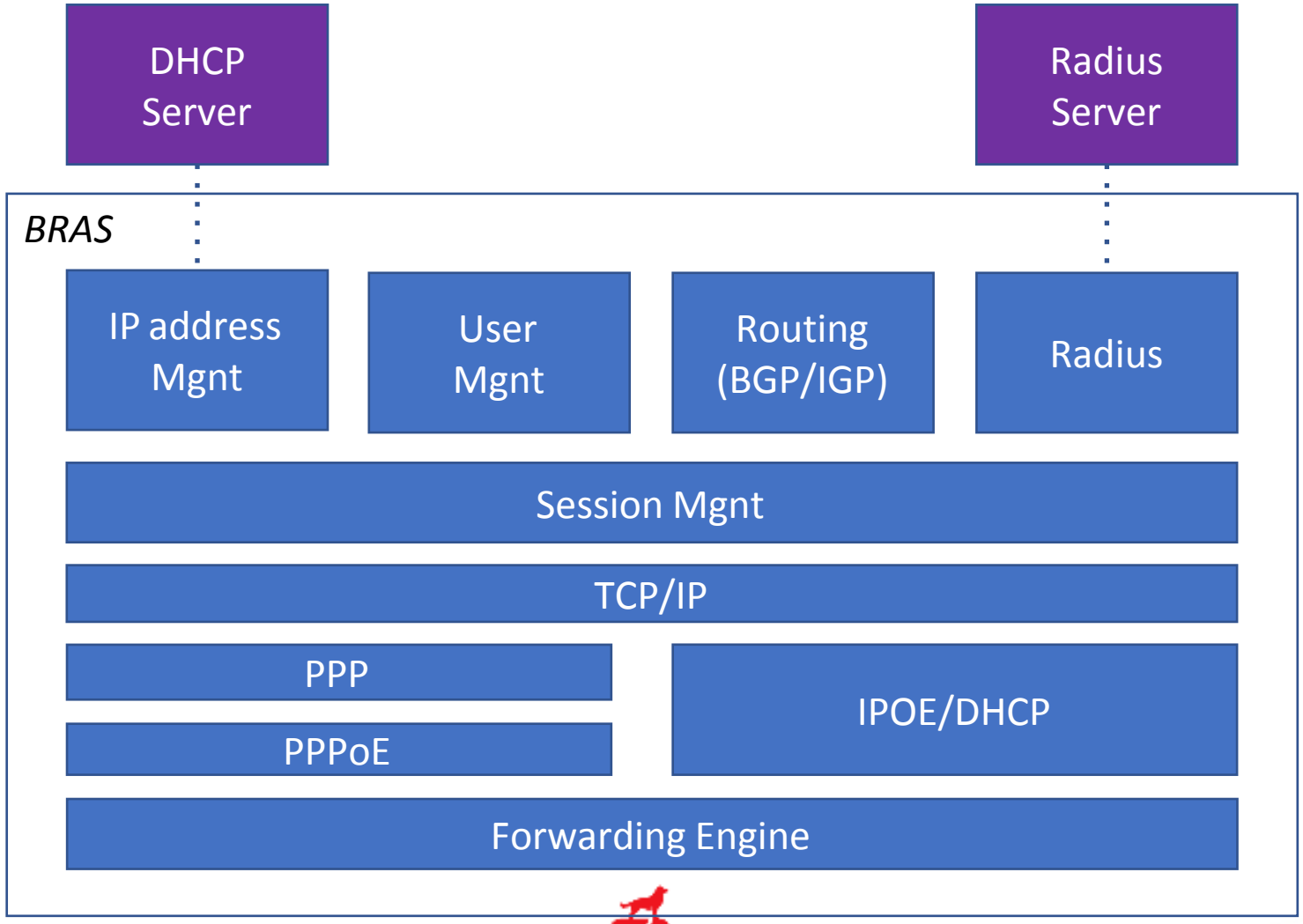
IP Network of China Telecom

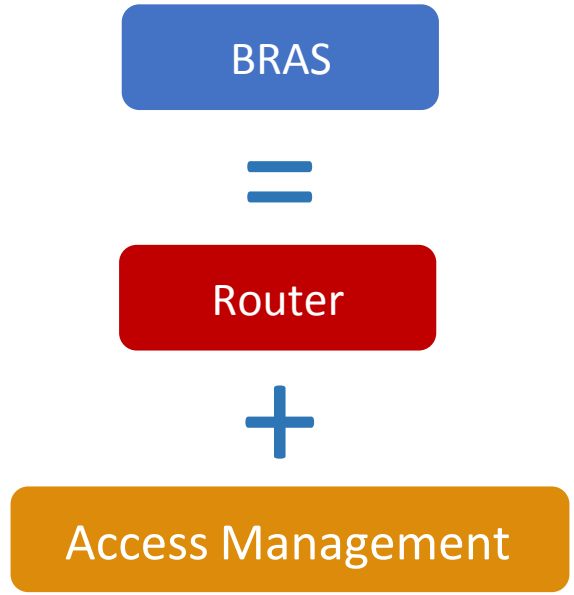
- **Chinanet (100 million BroadBand subscribers)**
  - ✓ IP Backbone for Internet Service
  - ✓ 150 cluster routers，2000 link
  - ✓ MAN interconnect，traffic aggregation
  - ✓ 300 MANs，Capacity over 30T
  - ✓ Over **10K BRAS hardware devices**

Biggest IP backbone in the world

.io Foundation

Traditional BRAS Architecture





Drawbacks:

1. Control Plane and Forwarding Plane are tightly coupled
2. Not flexible to scale

vBRAS Architecture

www.openbras.org.cn

vBRAS based on rack

vDHCP

vAAA

Microservice Bus

OpenBRAS Control Plane

Session discovery service

IPoE | PPP | PPPoE | L2TP

Data plane control service

Subscriber | Routing | ACL | QoS

Common service

ACE | libevent

Southbound Plugins

Openflow | NETCONF | RESTCONF

MANO

NFV-O

GVNFM

Cloud Management

OpenStack | Kubernetes

OpenBRAS Data Plane

Slow Path

Control Plane Agent

Classifier | Decap | ACL | Policer | Routing

Uplink

HQoS | Encap | ACL | Routing

Downlink

Supported Data Plane Options

Socket | VPP | DPDK | Commercial Software | Smart NIC

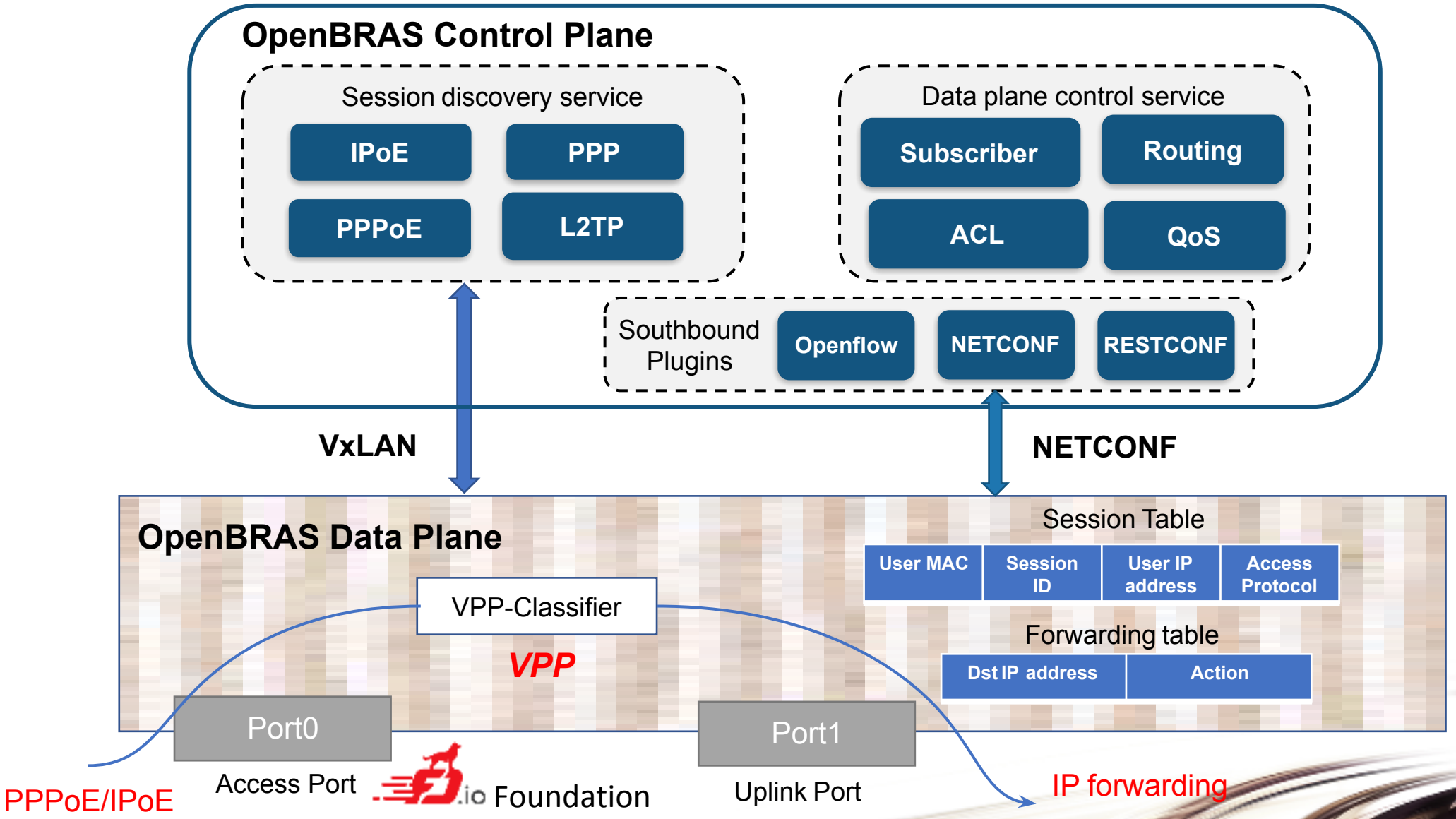
Legend

Control plane

Data plane

Foundation

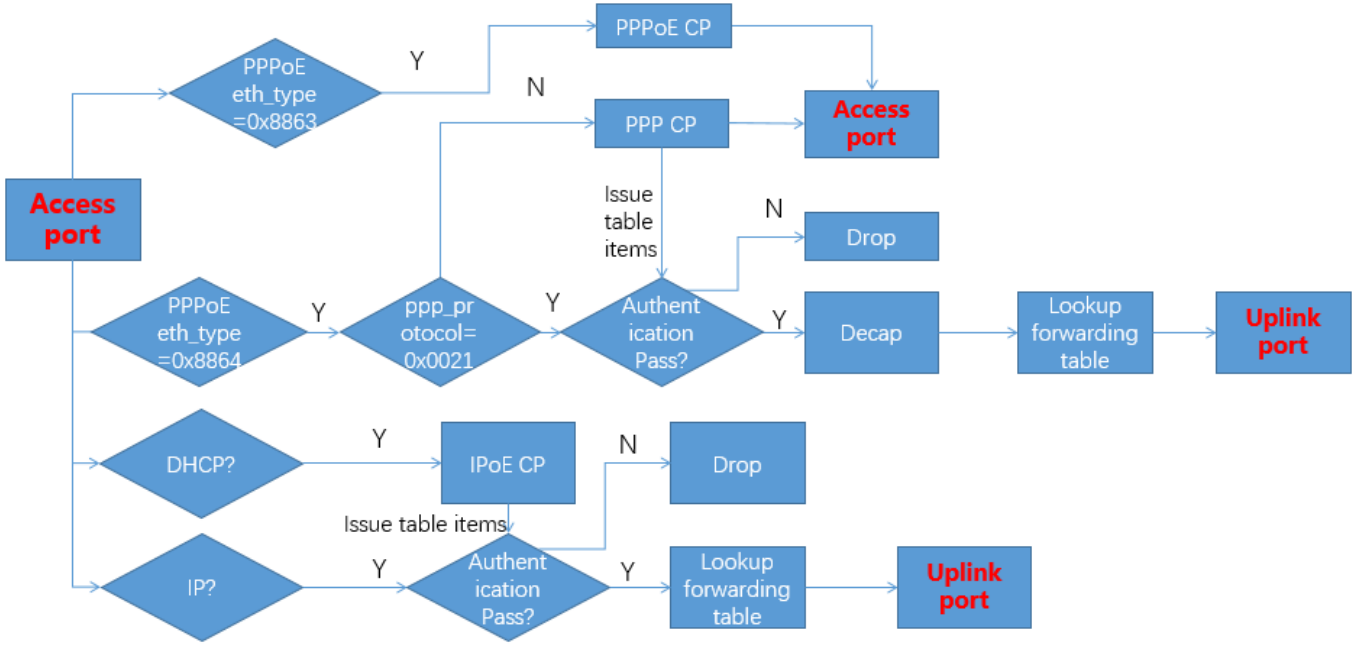
PPPoE/IPoE Processing of vBRAS



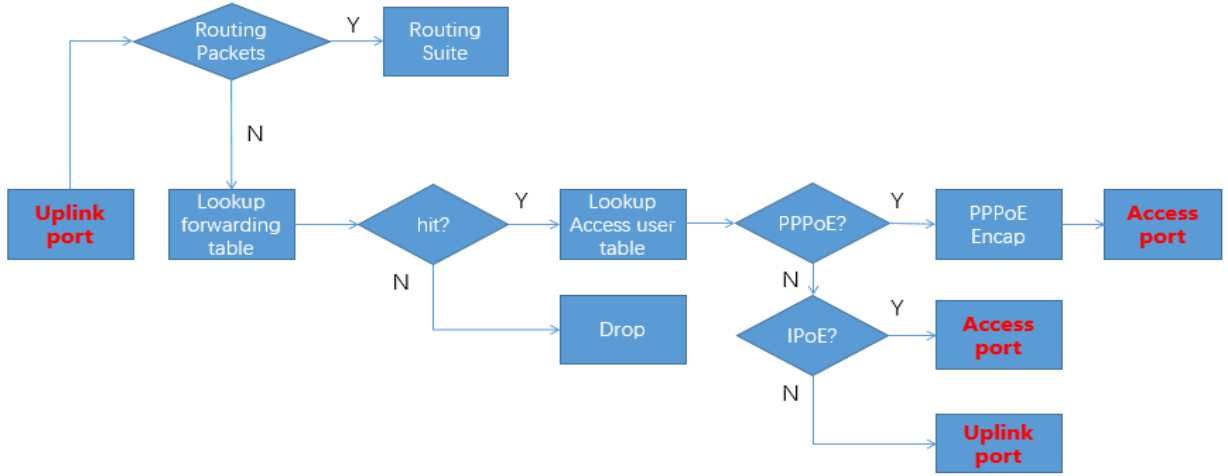
What does VPP need to do?

- Identify PPPoE Discovery and Session packets
  - With eth_type 0x8863 and 0x8864
- Identify PPP Packets with ppp_protocol 0x0021
  - Which are identified as IP packets
- Identify DHCP Packets with (udp_src_port, udp_dst_port) = (67, 68) or (68, 67)
- Identify Routing Packets (BGP/IGP)
- Encap or Decap PPPoE and PPP packet
  - Forward packets between access port and uplink port
- Look up table
  - Session table to determine whether the traffic is allowed to forward
  - Forwarding table to determine which port to forward the traffic

vBRAS Processing Flow and Pseudo Code

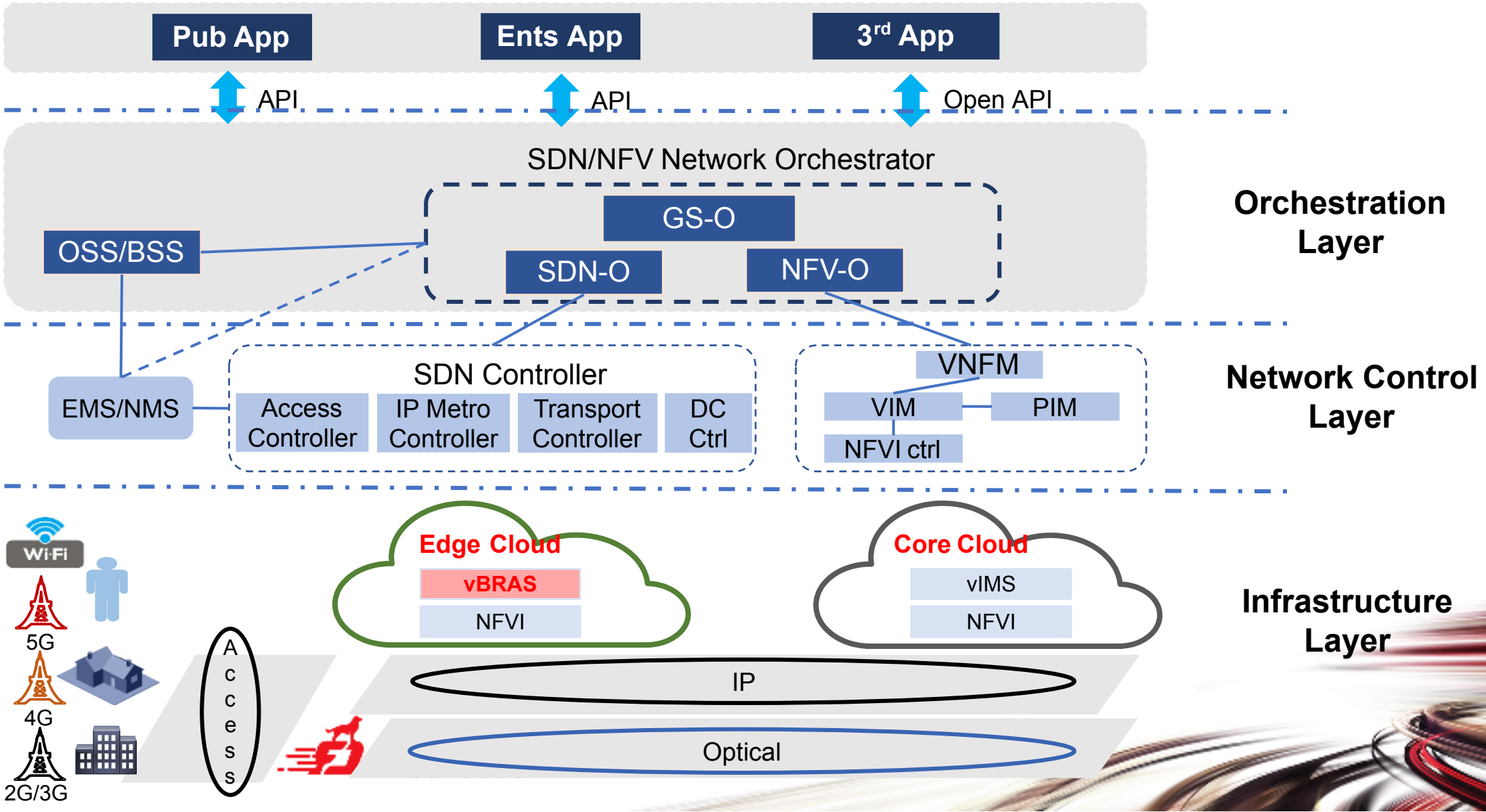



```
if packet from access_port
        if eth_type == 0x8863
                send to PPPoE_CP
        else if eth_type == 0x8864
                if ppp_type == 0x0021
                        if lookup_access_user_table(src_mac, src_ip, session id) == PPPoE
                                pppoe_decap & forword to forwarding_table(dst_ip)
                        else
                                drop
                else
                        send to PPP_CP
        else if eth_type == 0x0800
                if ip_protocol == 17 && (udp_src_port, udp_dst_port) = (67, 68) or (68,67)
                        send to IPOE_CP
                else if lookup_access_user_table(src_mac, src_ip) == IPoE
                        forword to forwarding_table(dst_ip)
                else
                        drop
        else
                drop
```

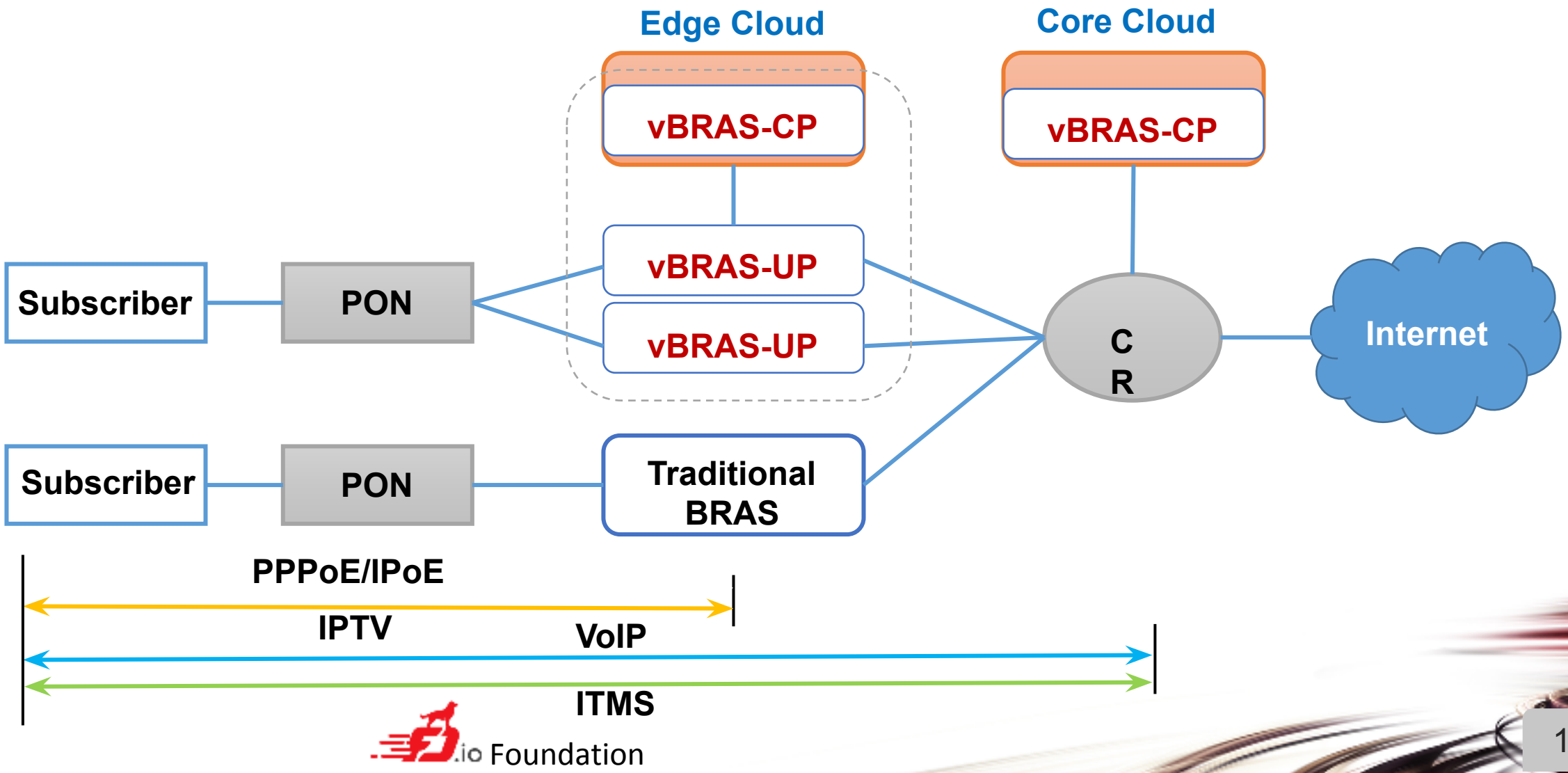



```
if packet from core_port
        if routing packets
                send to routing suite
        else if forwarding_table(dis_ip) !=NULL
                if lookup_access_user_table(dst_ip) == PPPoE
                        pppoe_encap & forward to forwarding_table(dst_ip)
                else if lookup_access_user_table(dst_ip) == IPoE
                        forword to forwarding_table(dst_ip)
                else
                        forword to forwarding_table(dst_ip)
        else
                drop
```

Where does vBRAS locate in China Telecom ?



Possible Deployment of vBRAS



OpenBRAS Promotes vBRAS Deployment

Project: Dedicated to build an open source vBRAS system, providing reference design

www.openbras.org.cn

China Telecom vBRAS Technical white Paper

- Include industry trends, **key elements**, **use cases**, and **solutions** of vBRAS

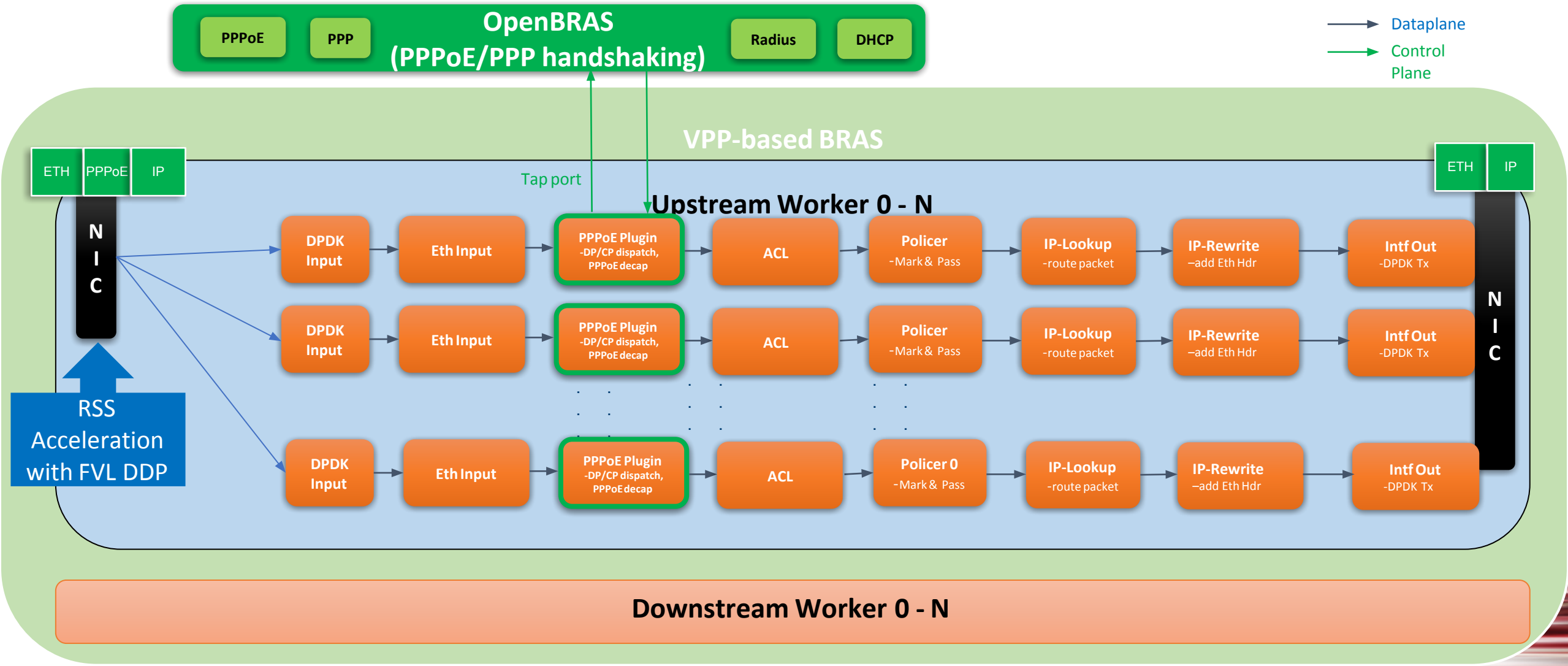
vBRAS CO Deployment Reference Solution

- Redesign **CO based on NFVI**, considering room area, floor bearing, power supply, cooling system
- Provide some typical device specifications

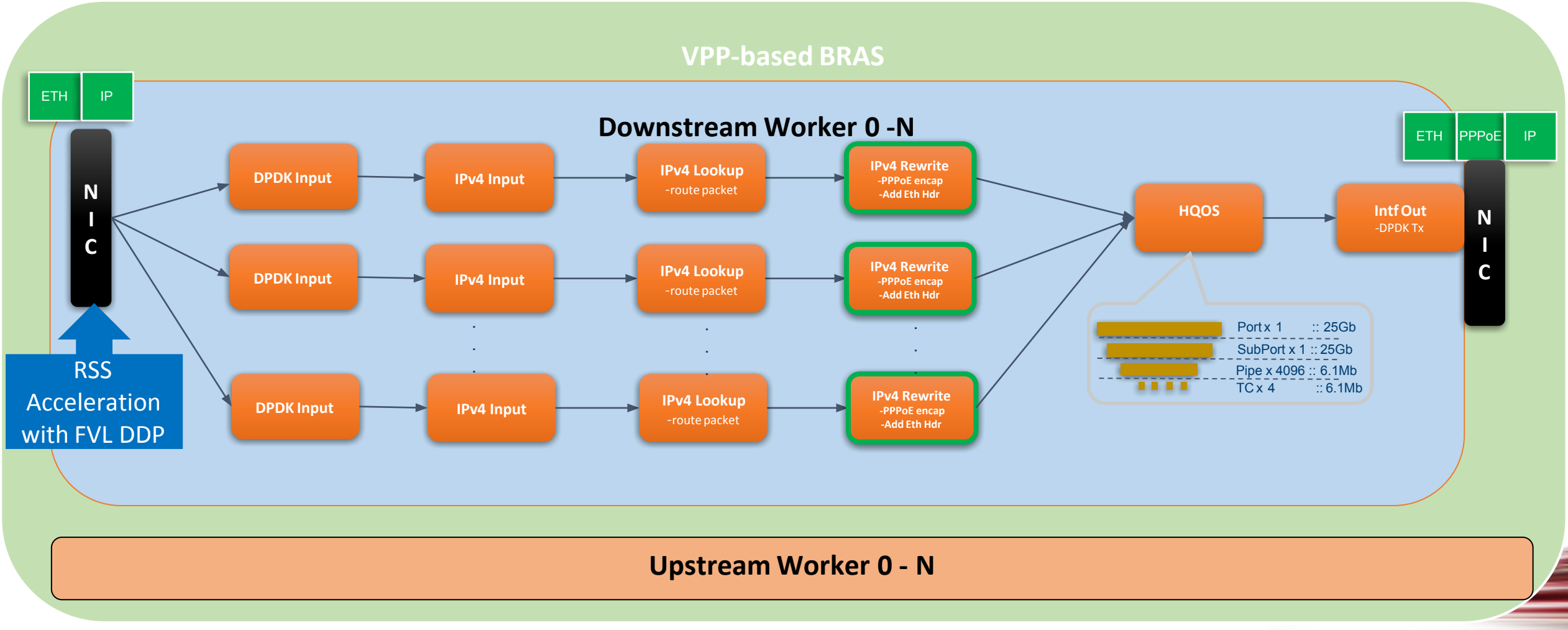
vBRAS offloading Solution

- vBRAS offloading solution based on **Smart NIC**
- Smart NIC: jobs like **PPPoE/PPP En/Decap**, **traffic forwarding**

Upstream Processing

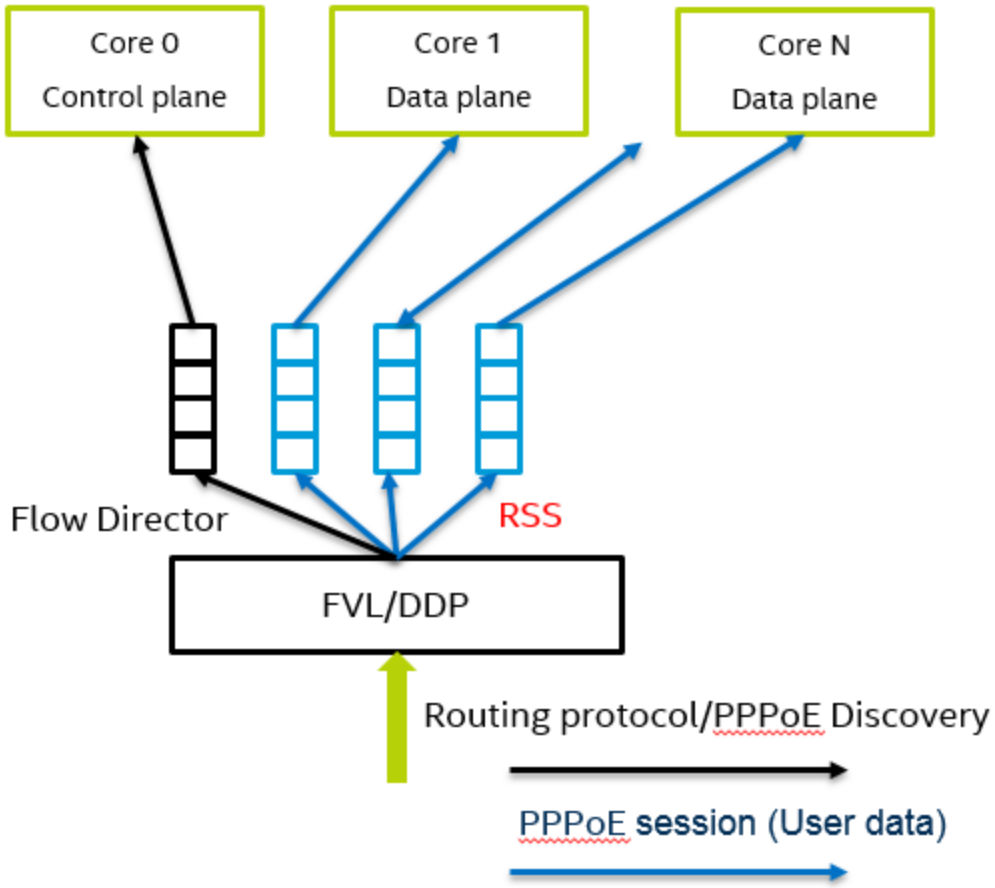


Downstream Processing

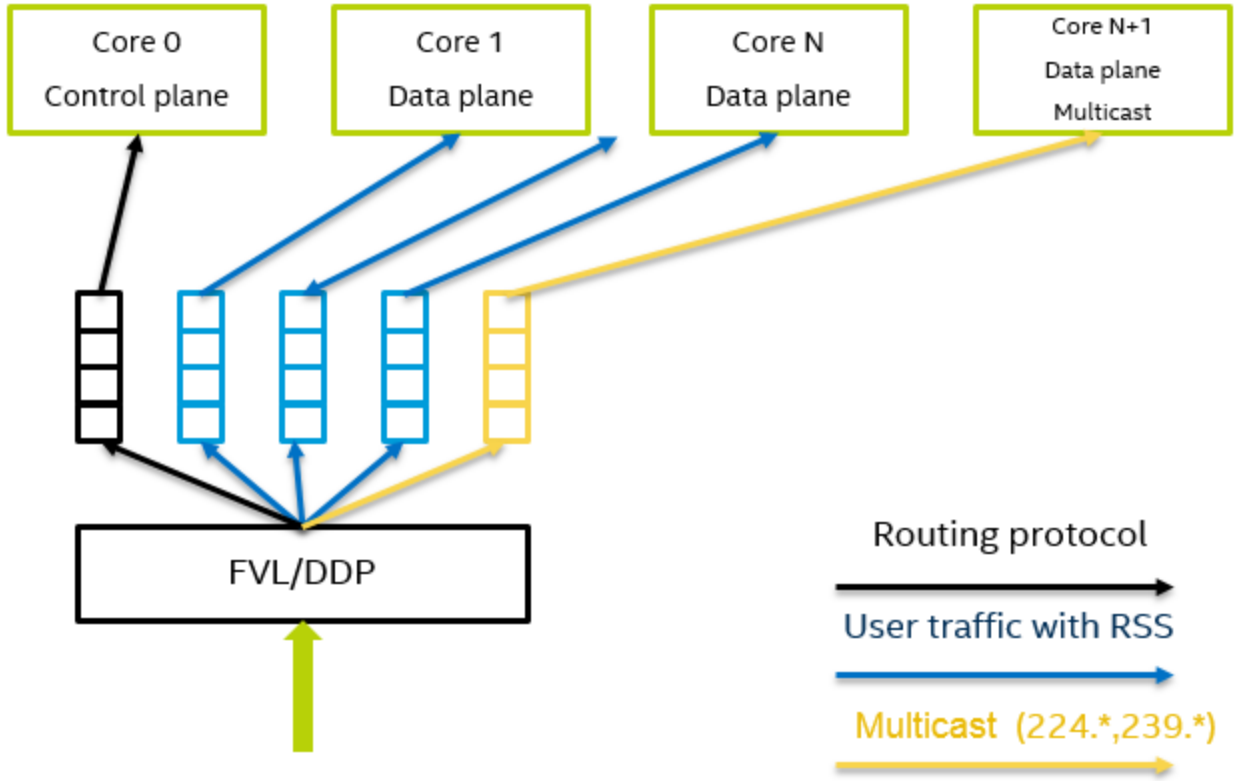


Distribute traffic on NIC

Upstream Traffic



Downstream Traffic



.io Foundation

Distrute traffic evenly

```
Thread 2 vpp_wk_1 (lcore 2)
Time 1.8, average vectors/node 255.99, last 128 main loops 20.00 per node 256.00
  vector rates in 3.9619e6, out 3.9619e6, drop 0.0000e0, punt 0.0000e0
             Name                State         Calls          Vectors        Suspends         Clocks       Vectors/Call
FortyGigabitEthernet7/0/0-outp   active              27781          7111931               0          7.77e0          255.99
FortyGigabitEthernet7/0/0-tx     active              27781          7111931               0          8.14e1          255.99
dpdk-input                       polling             27781          7111936               0          5.42e1          256.00
ethernet-input                   active              27781          7111936               0          2.95e1          256.00
ip4-inacl                        active              27781          7111931               0          9.67e1          255.99
ip4-input                        active              27781          7111931               0          3.55e1          255.99
ip4-lookup                       active              27781          7111931               0          2.32e1          255.99
ip4-policer-classify             active              27781          7111931               0          1.27e2          255.99
ip4-rewrite                      active              27781          7111931               0          2.15e1          255.99
pppoe-input                      active              27781          7111936               0          7.78e1          256.00
pppoe-tap-dispatch               active                  5                5               0          5.27e2            1.00
tuntap-tx                        active                  5                5               0          6.30e3            1.00
---------------
Thread 3 vpp_wk_2 (lcore 3)
Time 1.8, average vectors/node 255.99, last 128 main loops 20.00 per node 256.00
  vector rates in 3.9323e6, out 3.9323e6, drop 0.0000e0, punt 0.0000e0
             Name                State         Calls          Vectors        Suspends         Clocks       Vectors/Call
FortyGigabitEthernet7/0/0-outp   active              27574          7058942               0          7.58e0          255.99
FortyGigabitEthernet7/0/0-tx     active              27574          7058942               0          7.92e1          255.99
dpdk-input                       polling             27574          7058944               0          5.35e1          256.00
ethernet-input                   active              27574          7058944               0          2.97e1          256.00
ip4-inacl                        active              27574          7058942               0          9.96e1          255.99
ip4-input                        active              27574          7058942               0          3.43e1          255.99
ip4-lookup                       active              27574          7058942               0          2.37e1          255.99
ip4-policer-classify             active              27574          7058942               0          1.27e2          255.99
ip4-rewrite                      active              27574          7058942               0          2.39e1          255.99
pppoe-input                      active              27574          7058944               0          8.03e1          256.00
pppoe-tap-dispatch               active                  2                2               0          7.18e2            1.00
tuntap-tx                        active                  2                2               0          6.81e3            1.00
---------------
Thread 4 vpp_wk_3 (lcore 4)
Time 1.8, average vectors/node 255.99, last 128 main loops 20.00 per node 256.00
  vector rates in 4.0203e6, out 4.0203e6, drop 0.0000e0, punt 0.0000e0
             Name                State         Calls          Vectors        Suspends         Clocks       Vectors/Call
FortyGigabitEthernet7/0/0-outp   active              28191          7216894               0          6.81e0          255.99
FortyGigabitEthernet7/0/0-tx     active              28191          7216894               0          8.02e1          255.99
dpdk-input                       polling             28191          7216896               0          5.44e1          256.00
ethernet-input                   active              28191          7216896               0          2.96e1          256.00
ip4-inacl                        active              28191          7216894               0          9.72e1          255.99
ip4-input                        active              28191          7216894               0          3.34e1          255.99
ip4-lookup                       active              28191          7216894               0          2.28e1          255.99
ip4-policer-classify             active              28191          7216894               0          1.24e2          255.99
ip4-rewrite                      active              28191          7216894               0          2.14e1          255.99
pppoe-input                      active              28191          7216896               0          7.58e1          256.00
pppoe-tap-dispatch               active                  2                2               0          6.51e2            1.00
tuntap-tx                        active                  2                2               0          5.96e3            1.00
---------------
```

Test setup:

- 64K PPPoE sessions
- 4 cores for upstream
- 4 cores for downstream

PPPoE Plugin - Discovery and Session Setup


```
Packet 2

00:00:50:365990: dpdk-input
  TenGigabitEthernet5/0/0 rx queue 0
  buffer 0x1a88dd5: current data 0, length 60, free-list 0, clone-coun
  PKT MBUF: port 0, nb_segs 1, pkt_len 60
    buf_len 2176, data_len 60, ol_flags 0x180, data_off 128, phys_addr
    packet_type 0x0
    Packet Offload Flags
      PKT_RX_IP_CKSUM_GOOD (0x0080) IP cksum of RX pkt. is valid
      PKT_RX_L4_CKSUM_GOOD (0x0100) L4 cksum of RX pkt. is valid
  PPPOE_DISCOVERY: 00:11:01:00:00:01 -> ff:ff:ff:ff:ff:ff
00:00:50:366001: ethernet-input
  PPPOE_DISCOVERY: 00:11:01:00:00:01 -> ff:ff:ff:ff:ff:ff
00:00:50:366008: pppoe-tap-dispatch
  PPPoE dispatch from sw_if_index 2 next 1 error 0
```

- Dispatch packets between client and OpenBRAS.
- Learn PPPoE FIB.
- Create PPPoE session.
- Create downstream route entry automatically.

PPPoE Plugin – Upstream and Downstream

```
Packet 3

00:01:43:192310: dpdk-input
  TenGigabitEthernet5/0/0 rx queue 0
  buffer 0x1a88dae: current data 0, length 68, free-list 0, clone-count 0, totlen-
  PKT MBUF: port 0, nb_segs 1, pkt_len 68
    buf_len 2176, data_len 68, ol_flags 0x180, data_off 128, phys_addr 0x76632a80
    packet_type 0x0
    Packet Offload Flags
      PKT_RX_IP_CKSUM_GOOD (0x0080) IP cksum of RX pkt. is valid
      PKT_RX_L4_CKSUM_GOOD (0x0100) L4 cksum of RX pkt. is valid
  PPPOE_SESSION: 00:11:01:00:00:01 -> 90:e2:ba:48:7a:80
00:01:43:192318: ethernet-input
  PPPOE_SESSION: 00:11:01:00:00:01 -> 90:e2:ba:48:7a:80
00:01:43:192325: pppoe-input
  PPPoE decap from pppoe_session0 session_id 1 next 1 error 0
00:01:43:192329: ip4-input
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 63, length 46, checksum 0xb069
    fragment id 0x0000
00:01:43:192332: ip4-lookup
  fib 0 dpo-idx 2 flow hash: 0x00000000
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 63, length 46, checksum 0xb069
    fragment id 0x0000
00:01:43:192340: ip4-rewrite
  tx_sw_if_index 3 dpo-idx 2 : ipv4 via 100.1.1.100 TenGigabitEthernet5/0/1: 0000
  00000000: 00000000000190e2ba487a8108004500002e000000003effb169640101026401
  00000020: 0164000102030405060708090a0b0c0d0e0f10111213141516171819
00:01:43:192342: TenGigabitEthernet5/0/1-output
  TenGigabitEthernet5/0/1
  IP4: 90:e2:ba:48:7a:81 -> 00:00:00:00:00:01
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 62, length 46, checksum 0xb169
    fragment id 0x0000
00:01:43:192344: TenGigabitEthernet5/0/1-tx
  TenGigabitEthernet5/0/1 tx queue 0
  buffer 0x1a88dae: current data 8, length 60, free-list 0, clone-count 0, totlen-
  IP4: 90:e2:ba:48:7a:81 -> 00:00:00:00:00:01
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 62, length 46, checksum 0xb169
    fragment id 0x0000
```

```
Packet 1

00:01:23:344389: dpdk-input
  TenGigabitEthernet5/0/1 rx queue 0
  buffer 0xde8f56: current data 14, length 46, free-list 0, clone-count 0, totlen-nifb 0, trace 0x0
  PKT MBUF: port 1, nb_segs 1, pkt_len 60
    buf_len 2176, data_len 60, ol_flags 0x180, data_off 128, phys_addr 0x1b839480
    packet_type 0x0
    Packet Offload Flags
      PKT_RX_IP_CKSUM_GOOD (0x0080) IP cksum of RX pkt. is valid
      PKT_RX_L4_CKSUM_GOOD (0x0100) L4 cksum of RX pkt. is valid
  IP4: 90:e2:ba:48:7a:74 -> 90:e2:ba:48:7a:81
  RESERVED: 100.1.1.100 -> 100.1.1.2
    tos 0x00, ttl 64, length 46, checksum 0xaf69
    fragment id 0x0000
00:01:23:344394: ip4-input-no-checksum
  RESERVED: 100.1.1.100 -> 100.1.1.2
    tos 0x00, ttl 64, length 46, checksum 0xaf69
    fragment id 0x0000
00:01:23:344396: ip4-lookup
  fib 0 dpo-idx 3 flow hash: 0x00000000
  RESERVED: 100.1.1.100 -> 100.1.1.2
    tos 0x00, ttl 64, length 46, checksum 0xaf69
    fragment id 0x0000
00:01:23:344397: ip4-midchain
    CPNX: 136.100.17.0 -> 0.1.0.0
      version 0, header length 0
      tos 0x11, ttl 186, length 256, checksum 0x7a80 (should be 0xffff)
      fragment id 0x002e offset 34576, flags
00:01:23:344403: adj-midchain-tx
  adj-midchain:[3]:ipv4 via 0.0.0.0  pppoe_session0: 00110100000190e2ba487a8088641100000100000021
  stacked-on:
    [@3]: TenGigabitEthernet5/0/0-dpo:
00:01:23:344404: TenGigabitEthernet5/0/0-output
  pppoe_session0
  PPPOE_SESSION: 90:e2:ba:48:7a:80 -> 00:11:01:00:00:2e
00:01:23:344406: TenGigabitEthernet5/0/0-tx
  TenGigabitEthernet5/0/0 tx queue 0
  buffer 0xde8f56: current data -8, length 68, free-list 0, clone-count 0, totlen-nifb 0, trace 0x0
  PPPOE_SESSION: 90:e2:ba:48:7a:80 -> 00:11:01:00:00:2e
```

ACL and Policer

ACL:

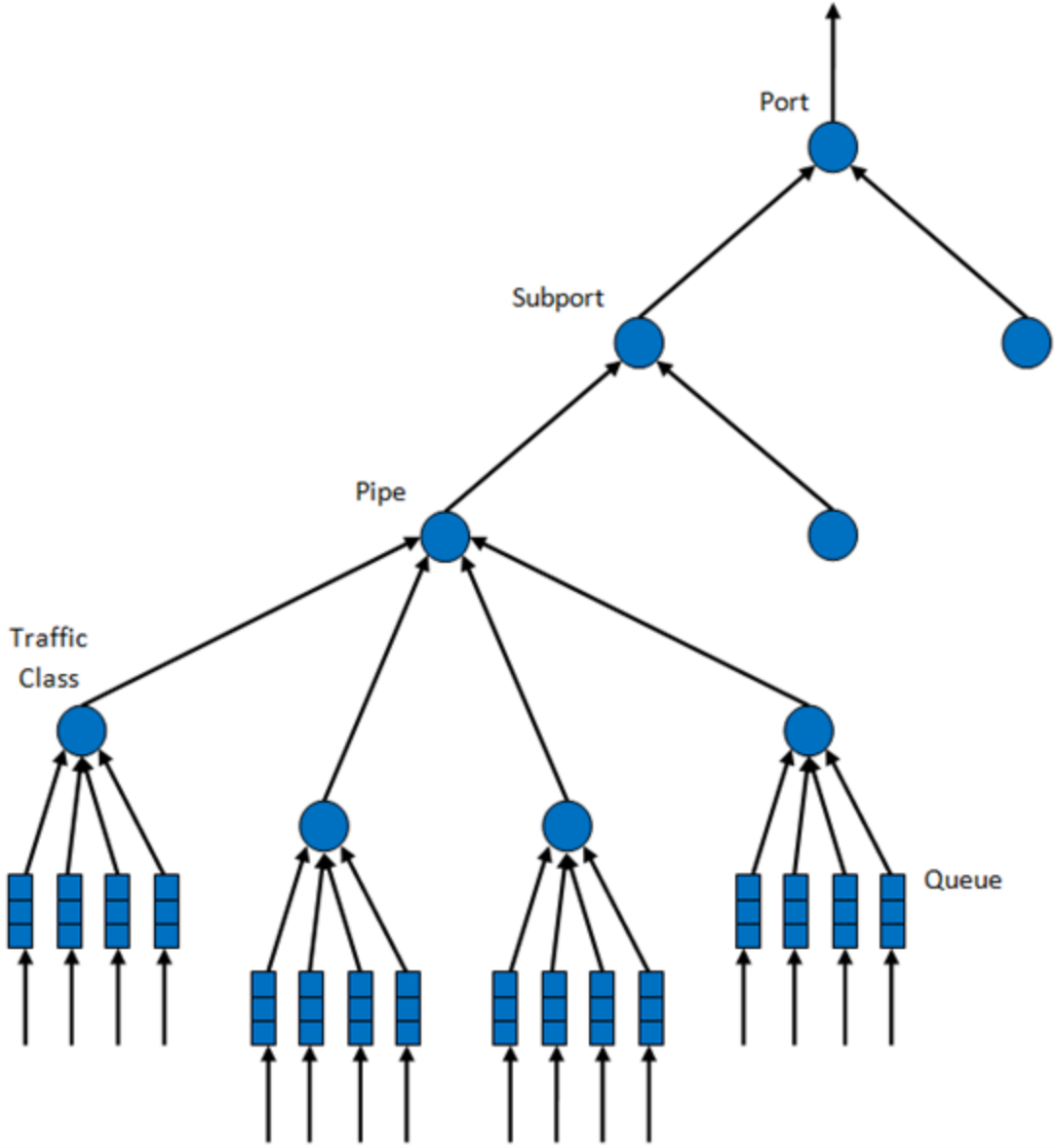
- One ACL rule for each subscriber.
- Configurable to match any field.

Policer:

- One policer rule for each subscriber.
- Configurable to match any field.
- Actions: conform/exceed/violate.

```
00:02:00:289251: dpdk-input
  TenGigabitEthernet5/0/0 rx queue 0
  buffer 0xde8eba: current data 0, length 68, free-list 0, clone-count 0, tot
  PKT MBUF: port 0, nb_segs 1, pkt_len 68
    buf_len 2176, data_len 68, ol_flags 0x180, data_off 128, phys_addr 0x1b83
    packet_type 0x0
    Packet Offload Flags
      PKT_RX_IP_CKSUM_GOOD (0x0080) IP cksum of RX pkt. is valid
      PKT_RX_L4_CKSUM_GOOD (0x0100) L4 cksum of RX pkt. is valid
  PPPOE_SESSION: 00:11:01:00:00:01 -> 90:e2:ba:48:7a:80
00:02:00:289256: ethernet-input
  PPPOE_SESSION: 00:11:01:00:00:01 -> 90:e2:ba:48:7a:80
00:02:00:289259: pppoe-input
  PPPoE decap from pppoe_session0 session_id 1 next 1 error 0
00:02:00:289261: ip4-input
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 63, length 46, checksum 0xb069
    fragment id 0x0000
00:02:00:289262: ip4-inacl
  INACL: sw_if_index 2, next_index 2, table 0, offset 192
00:02:00:289266: ip4-policer-classify
  POLICER_CLASSIFY: sw_if_index 2 next 1 table 1 offset 192 policer_index 0
00:02:00:289268: ip4-lookup
  fib 0 dpo-idx 2 flow hash: 0x00000000
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 63, length 46, checksum 0xb069
    fragment id 0x0000
00:02:00:289269: ip4-rewrite
  tx_sw_if_index 3 dpo-idx 2 : ipv4 via 100.1.1.100 TenGigabitEthernet5/0/1:
  00000000: 00000000000190e2ba487a8108004500002e000000003effb169640101026401
  00000020: 0164000102030405060708090a0b0c0d0e0f10111213141516171819
00:02:00:289271: TenGigabitEthernet5/0/1-output
  TenGigabitEthernet5/0/1
  IP4: 90:e2:ba:48:7a:81 -> 00:00:00:00:00:01
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 62, length 46, checksum 0xb169
    fragment id 0x0000
00:02:00:289271: TenGigabitEthernet5/0/1-tx
  TenGigabitEthernet5/0/1 tx queue 0
  buffer 0xde8eba: current data 8, length 60, free-list 0, clone-count 0, tot
  IP4: 90:e2:ba:48:7a:81 -> 00:00:00:00:00:01
  RESERVED: 100.1.1.2 -> 100.1.1.100
    tos 0x00, ttl 62, length 46, checksum 0xb169
    fragment id 0x0000
```

Hierarchical Qos (HQos based on DPDK) added to VPP



Level	Siblings per Parent	Description
Port	.	Output Ethernet port 1/10/40 GbE.
Subport	Configurable (default: 8)	A predefined group of users
Pipe	Configurable (default: 4K)	An individual user/subscriber
Traffic Class (TC)	4	A different traffic type with specific loss rate, delay and jitter.
Queue	4	Hosts packets from multiple connection of the same type belonging to the same user

Key Takeaway

http://fd.io/

- A solution with separated Control Plane and Data Plane.
- Implementation ready for Cloud Edge deployment.
- Control Plane leverages OpenBRAS, an open source project.
- Data Plane leverages VPP and DPDK.
- Distribute traffic from NIC to multiple cores evenly.