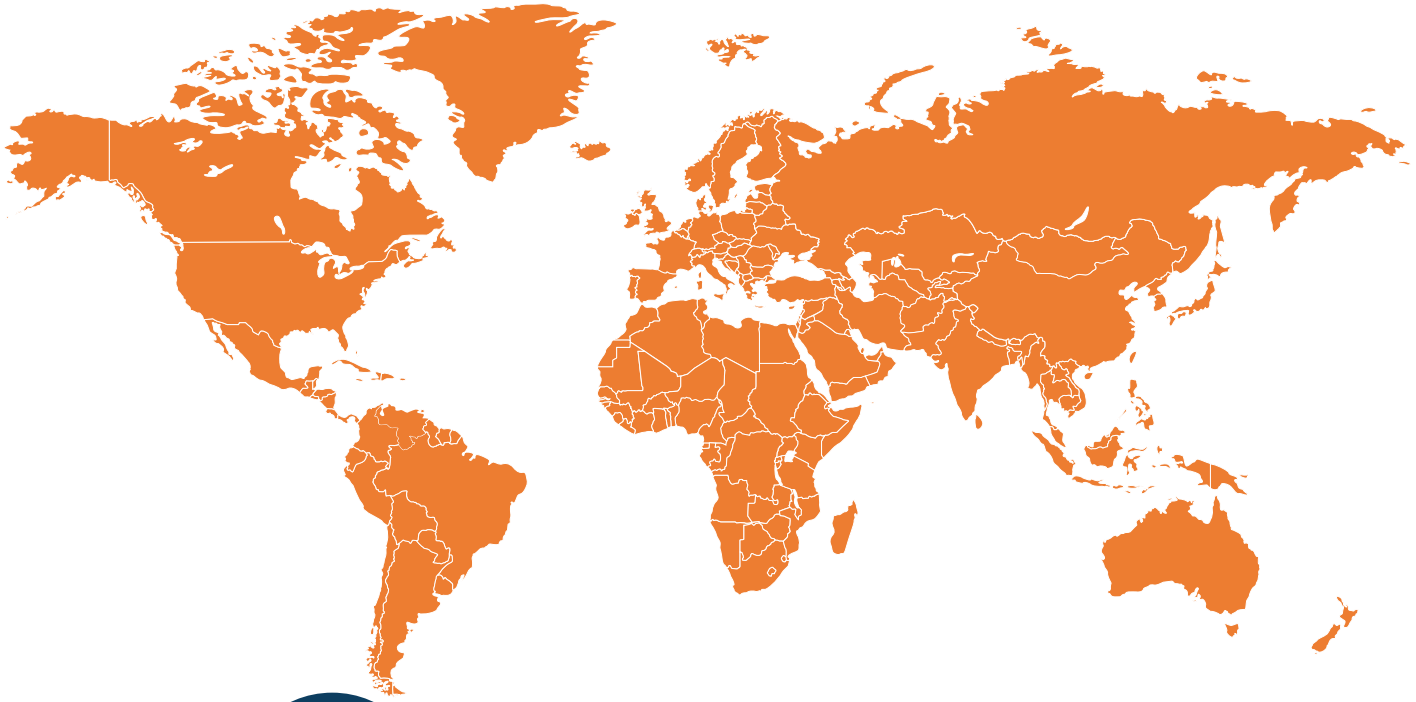
ISP Design: Separation of Network Functions

Modularity in design with commodity equipment for rapid scale while reducing CAPEX

IP Architechs the story

IP ArchiTechs is a global network engineering and design firm that covers a wide spectrum of environments. Our expertise spans service providers, datacenters, and enterprises.



GLOBAL

Offices in the US, Europe and South America

6

100+ Active Projects

1000s of clients across 6 continents

ArchiTechs
MANAGED SERVICES

iparchitechs.com

+1 (855) 645-7684

We design, build, and troubleshoot IPv4/IPv6 Networks

At our core, we are a team of network engineers and architects that consult at every stage in a network's lifecycle.

- Network Engineering
- Network Design & Architecture
- Network Audit & Documentation
- Managed Network Services

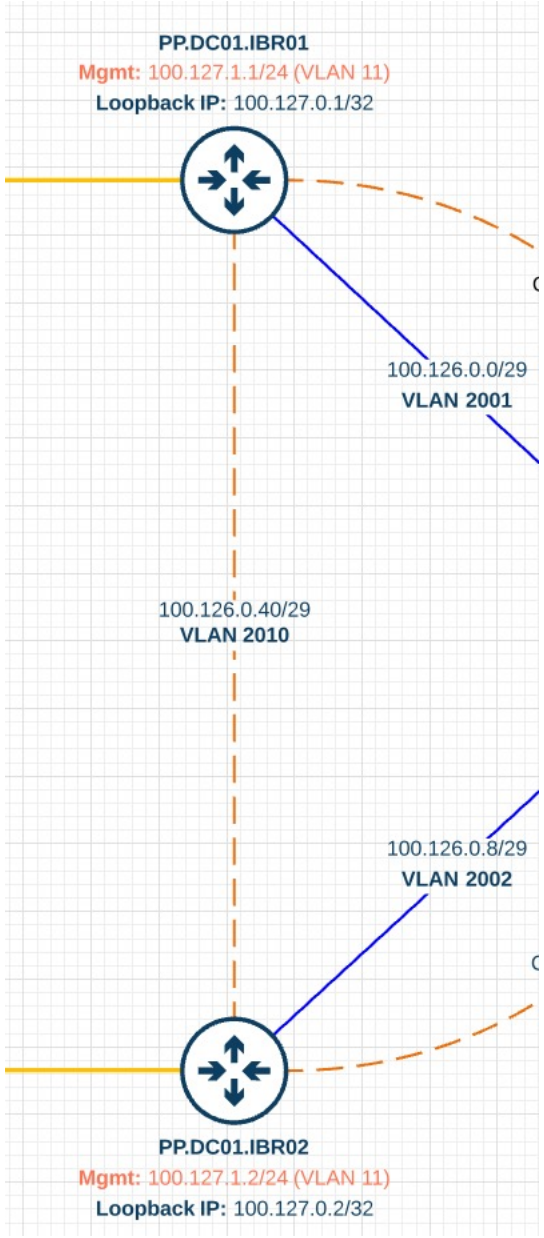
IPA | What we do?

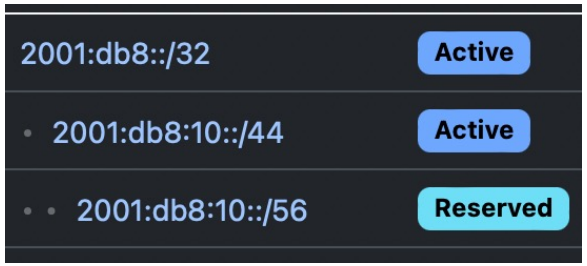
Here are some typical engagement contexts our clients come to us with.

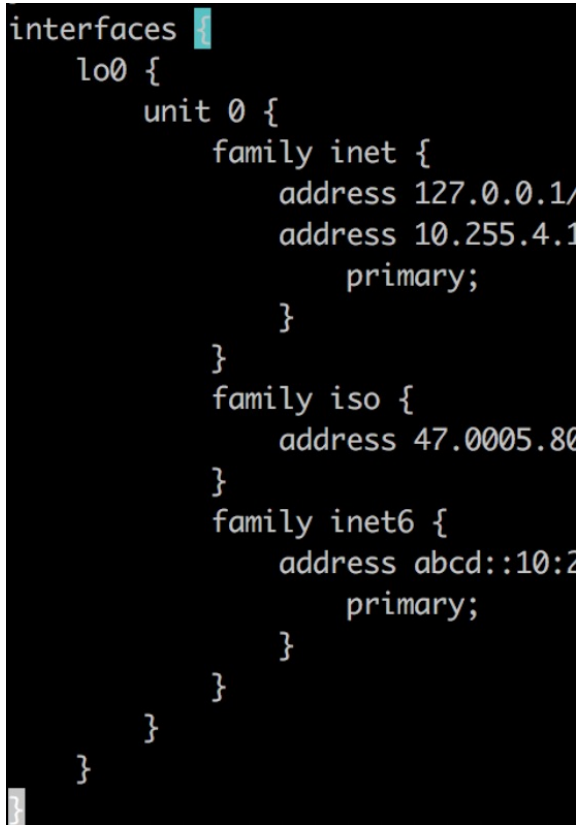
- Network Discovery/Audit – Consultative review of observations and findings
- Strategy consultation on network growth cycles. (Bridged -> Routed -> Basic Dynamic Routing -> and beyond)
- Systems/Software Integration and Automation
- Network migration planning, simulation, execution. We can take the driver seat and get things done that you otherwise don't have time for.
- Break/Fix Troubleshooting and Incident Response

We'll work with you to identify the best path forward, considering business goals. We can help strategize, plan, configure, and implement.

The operational tempo of our team has us uniquely trained to plan and lead executions of network migrations at any size.







Understanding Separation of Network Functions

01 What are network functions?

02 Design examples for WISP/FISP and Telco

03 Equipment and budget considerations

S1 | Network functions overview

What are network functions?

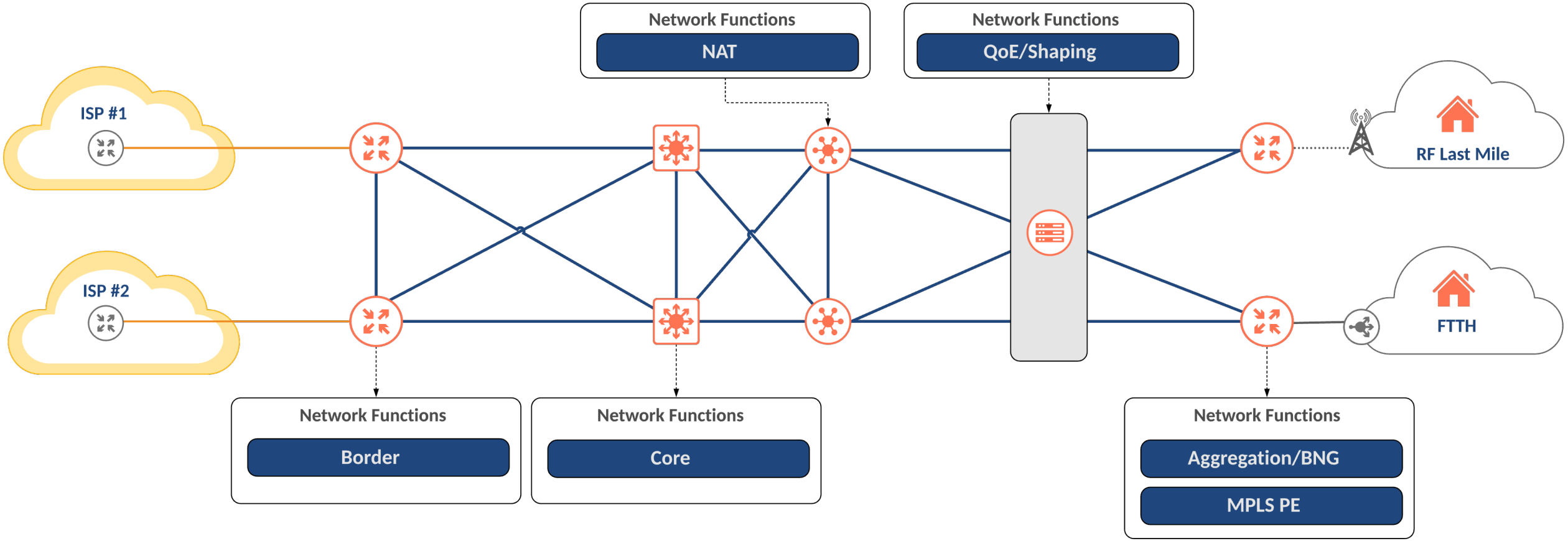
Section 1 | What are network functions?

01

- Overview
- Border / Core / Aggregation
- NAT and QoE
- Operational Support Functions

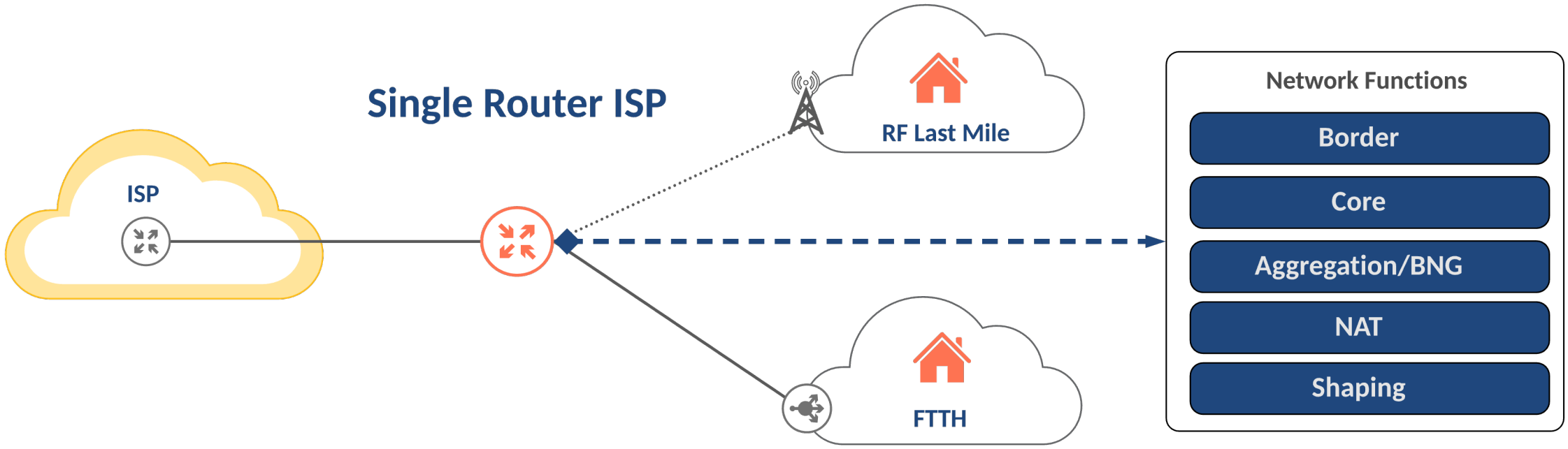
Overview of an ISP capable of 10,000+ subscribers with separated functions

Leveraging whitebox and commodity vendors for scale - At first glance, this may seem like a lot of gear, but as we go through the individual functions, the flexibility and scalability will become clearer. A decade ago this architecture would be unaffordable with vendors like Cisco and Juniper, but thanks to whitebox and commodity vendors like IP Infusion, Edge Core, MikroTik and UfiSpace, WISPs all over the world have production networks capable of 10s of thousands of subscribers using this design.



What are network functions?

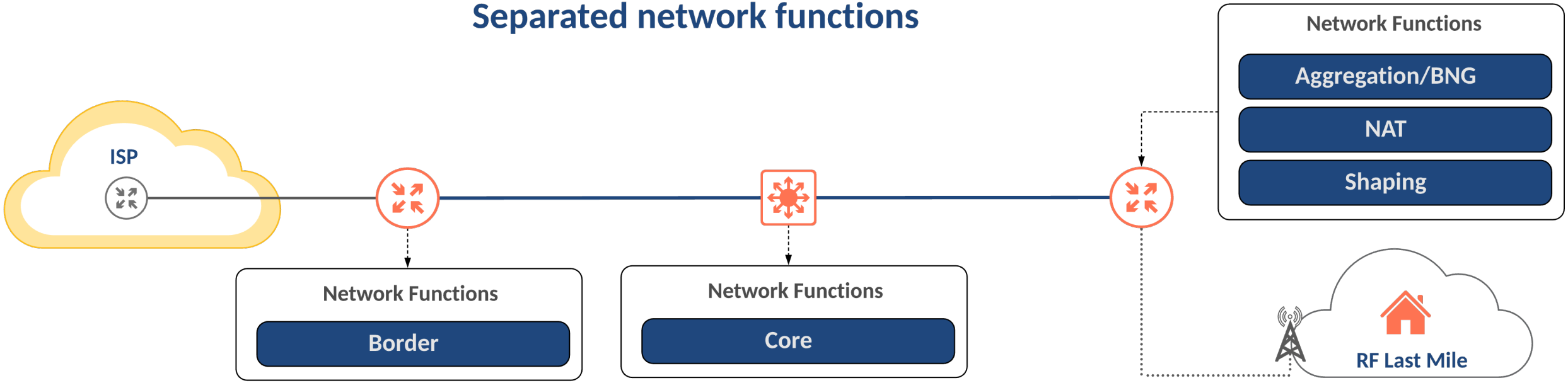
Network Functions - The major tasks in the data plane that must be performed by an L2/L3 network device to ensure smooth delivery of the Internet from the border of an ISP down to the subscriber last mile. Examples are border routers, core switches and aggregation routers.

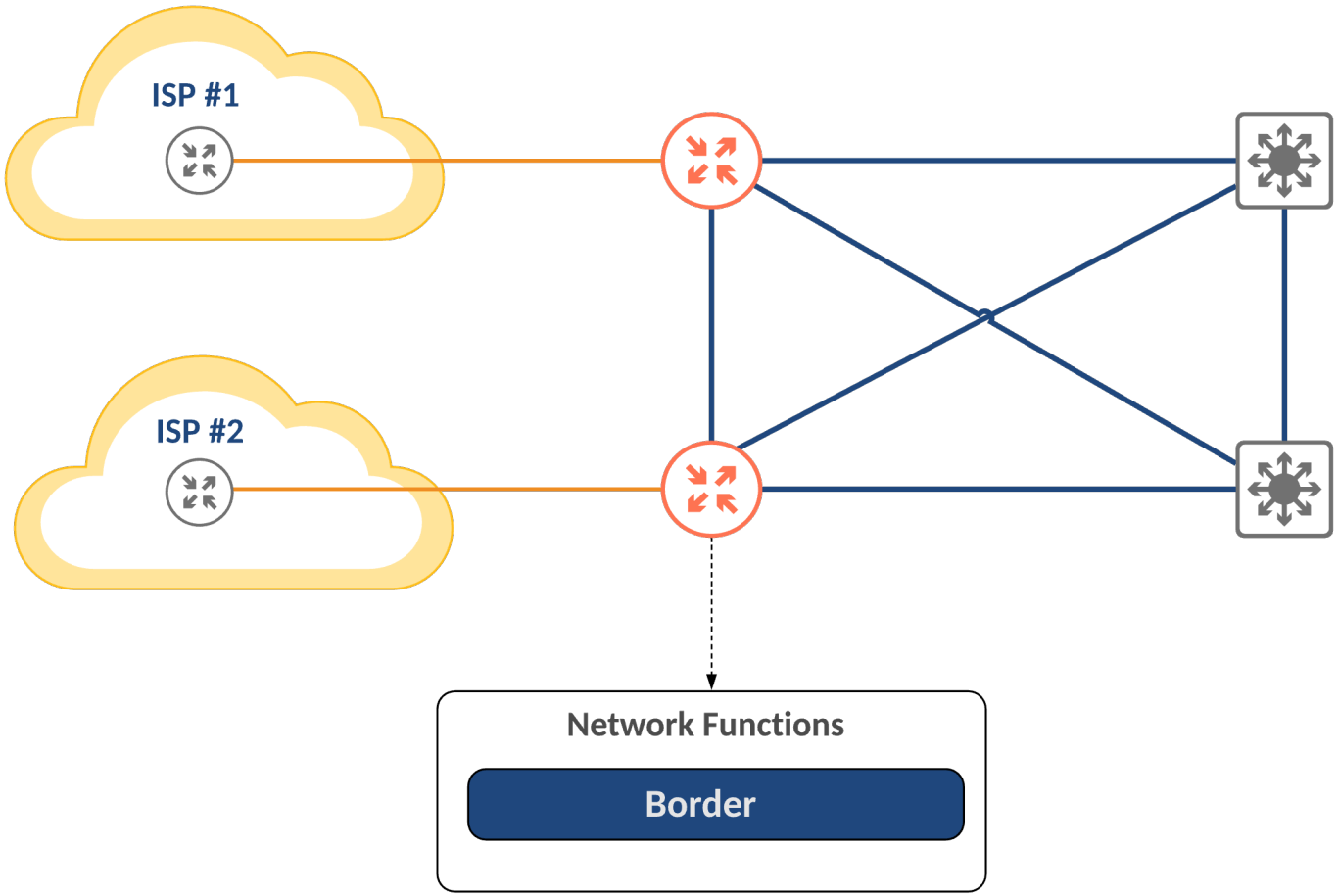


Why separate network functions?

It's tempting for new and even experienced ISPs to pile all of the functions into one router, switch or server - and then add another for "redundancy". This generally creates problems with complexity, failure domains and growth. Separating functions allows for network designs to be modular, repeatable and more scalable. Automation is easier because templating is easier. The end result is better uptime, lower opex, easier growth and lowered risk.

Separated network functions





What is the Border function?

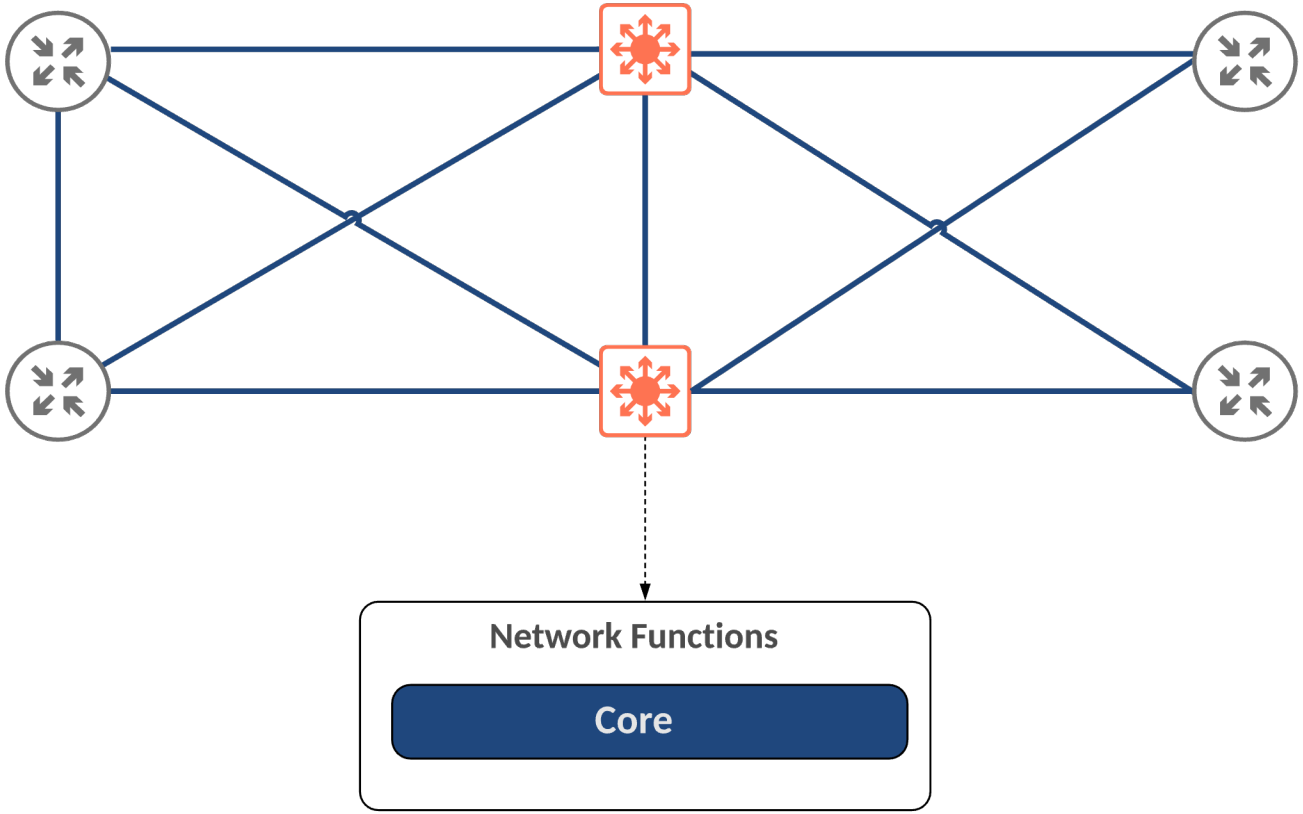
Border - Also called the edge, the border function terminates BGP peerings or DIA to connect to another ASN for internet transit.

Benefits - By separating this role:

- The network can scale Internet connections and peering routers up or down.

- Capacity and redundancy can be added with additional routers. Software upgrades are less impactful.

- Configuration is simplified and focused only on upstream connectivity.



What is the Core function?

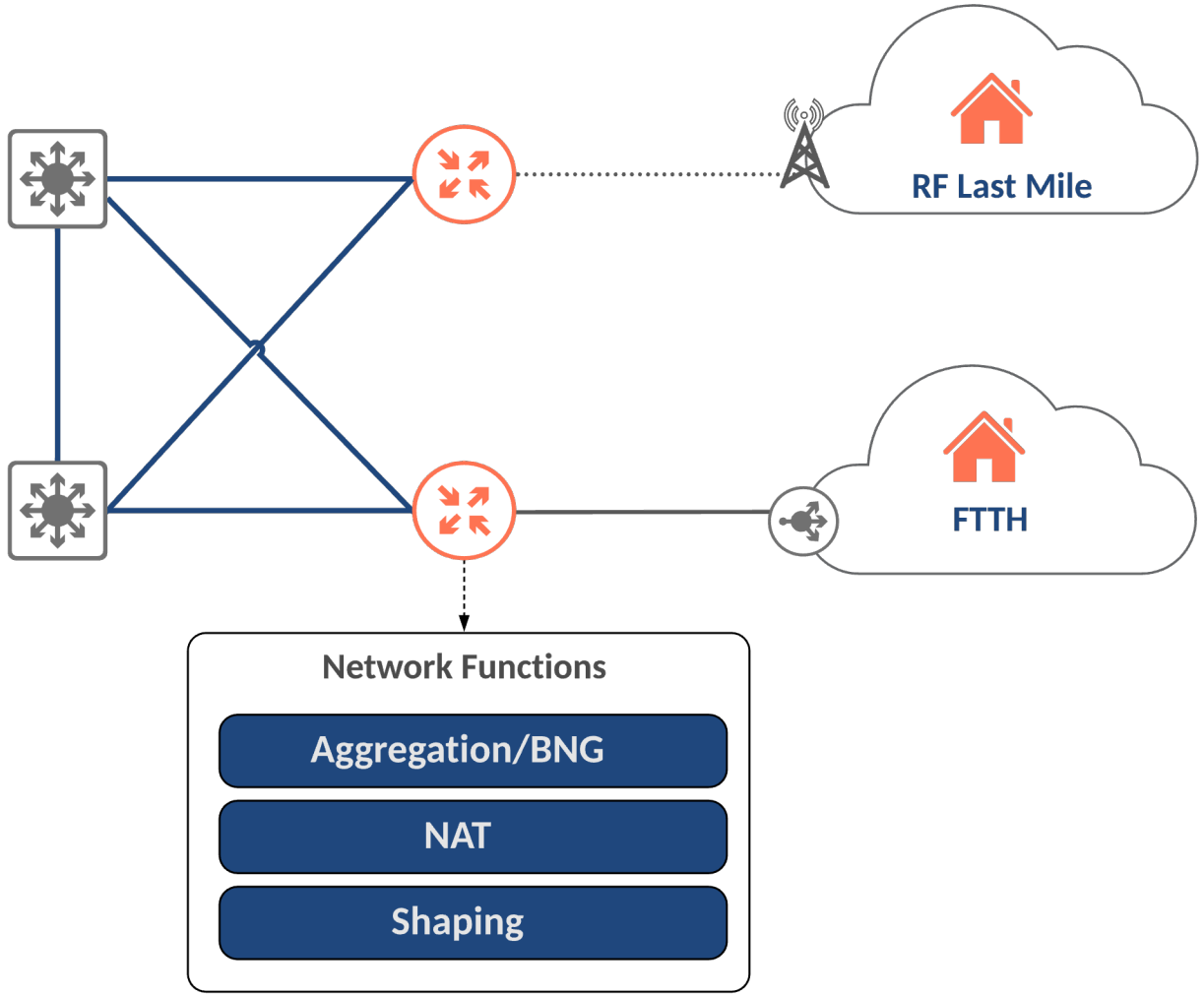
Core - The job of the network core is to connect all other devices and functions as simply as possible.

Benefits - By separating this role:

- The network can add routers, switches, firewalls, servers and storage as needed.

- Failover domains are separated between border and aggregation. A border router can fail and it won't affect subscriber facing services.

- Configuration is simplified and focused only on high speeds and port count.



What is the Aggregation function?

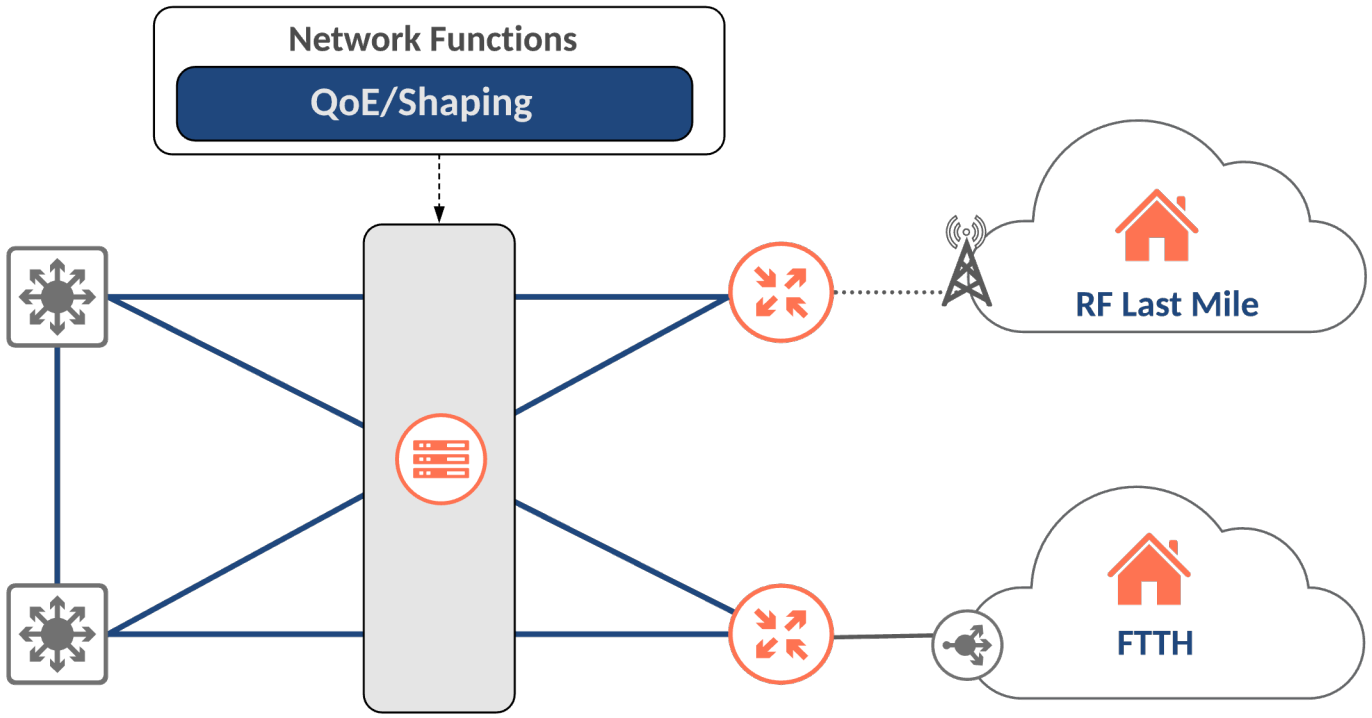
Aggregation - Also called a BNG or PE, these devices terminate services for subscribers and are often the L3 gateway for the last mile.

Benefits - By separating this role:

- Complexity in L2/L3, security and billing integration can be pushed to this layer and scaled horizontally via routers as needed for capacity.

- Combining other functions like NAT and Shaping are more practical in this layer.

- Specialized services like IPTV, SIP and Business Class offerings can be added as separate aggregation hardware if needed.



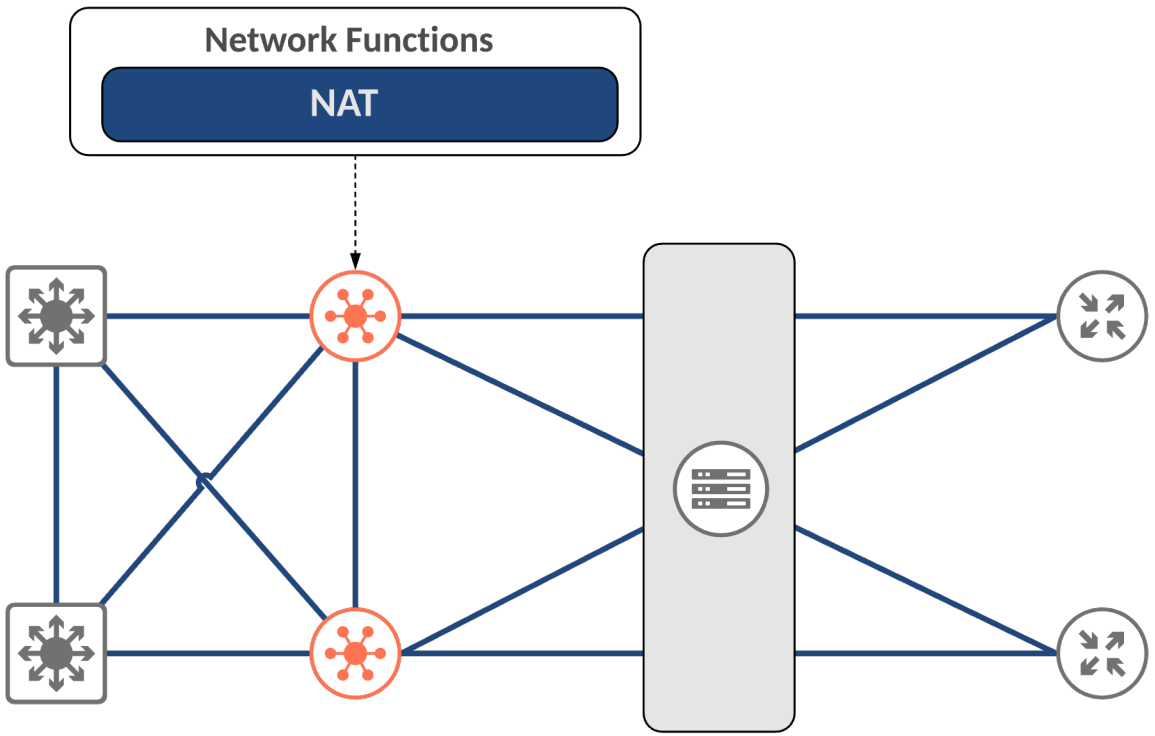
What is the QoE/Shaping function?

QoE/Shaping - Quality of Experience devices perform intelligent shaping by subscriber to ensure consistent performance. They can also provide analytics and performance data about network health. While it is possible to shape on the aggregation layer, QoE often requires a separate appliance thatsits between routers at L2.

Benefits - By placing QoE between core/agg:

- Intelligent shaping can be done without needing to add ports to support last mile additions.

- If NAT is required, separate routers can be used inline without a network redesign.



What is the NAT function?

Network Address Translation - NAT or CG-NAT is an IPv4 conservation function that is often combined with aggregation unless a QoE shaper is used.

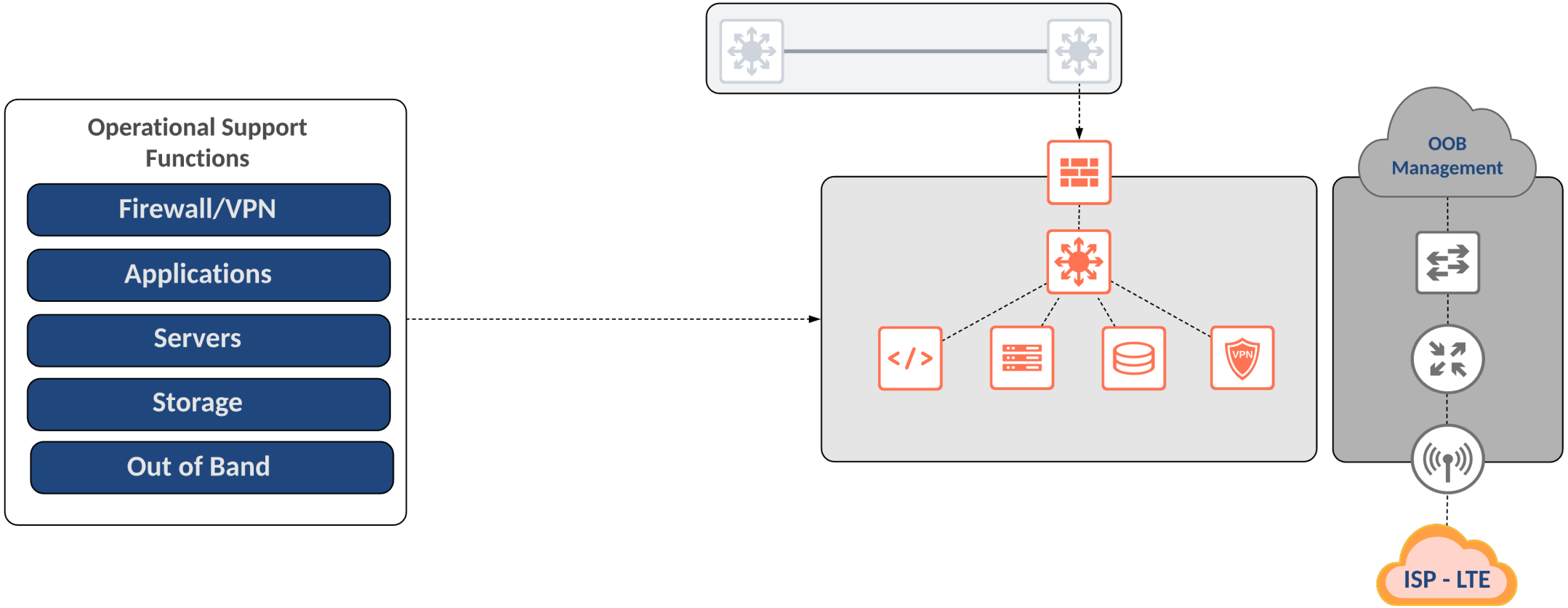
Benefits - By placing NAT between core/QoE:

- The QoE engine can see the original IP source of traffic before address translation. Otherwise this would have to be performed on the core which leads to bad design.

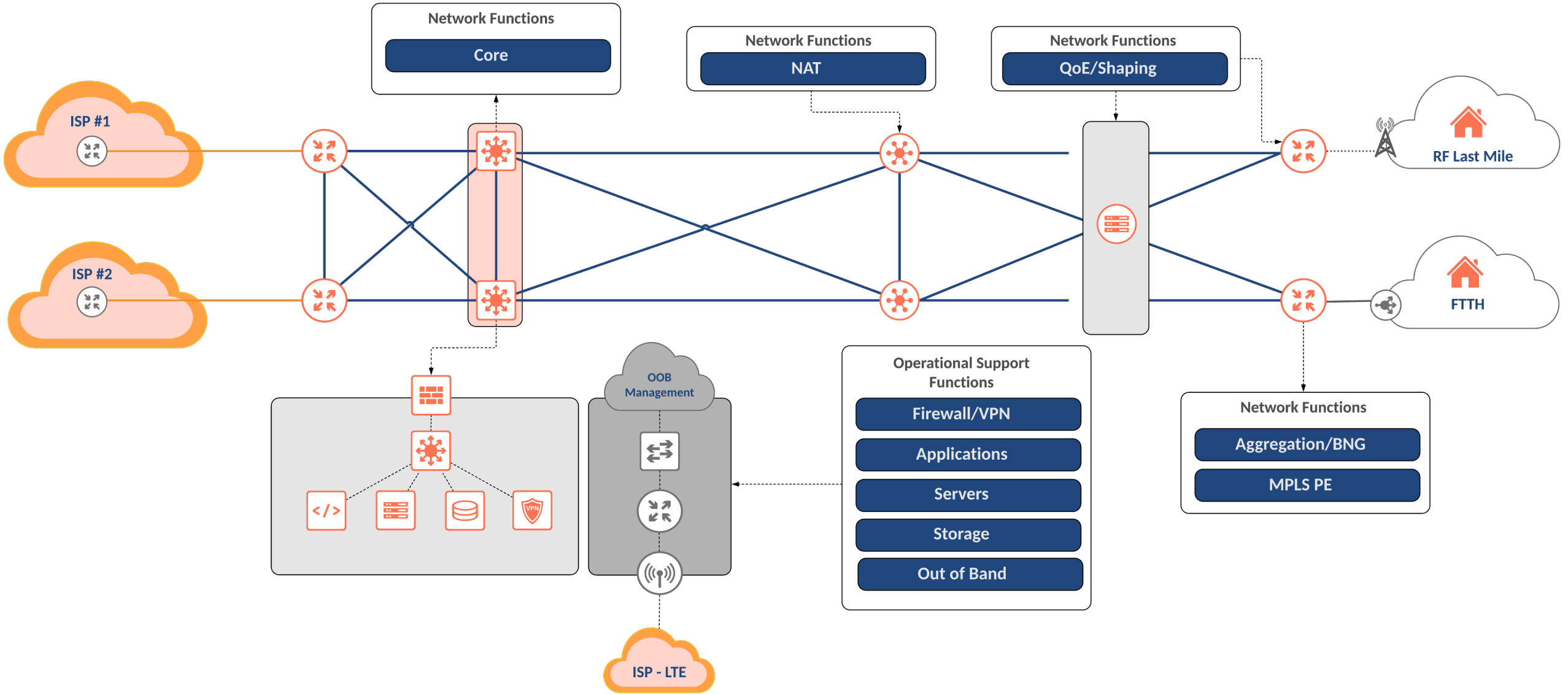
- Logging of state tables for CALEA compliance is an intensive task so creating dedicated routers for this task makes reporting easier.

What are Operational Support Functions?

Operational Support Functions - The major tasks in the control and management plane that must be performed by a device or service to facilitate and support the operation of network functions. Examples are DHCP, DNS, Applications/Servers, Billing Systems, Corporate VPNs and connectivity.

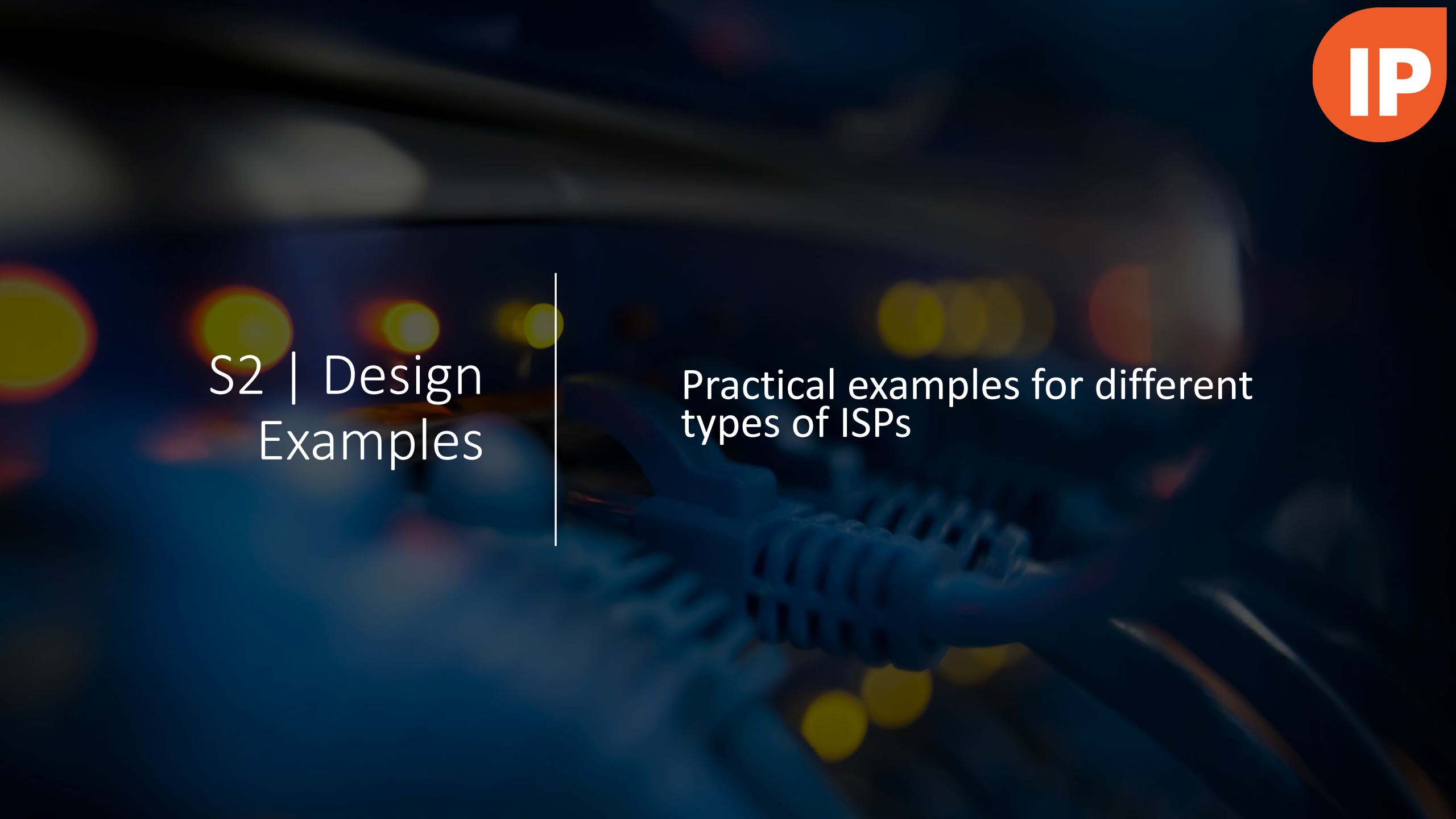


Putting it all together



Discussion: How far should I separate / how much can I collapse ?

- Different answers for different types of ISPs
  - **DC ISP** – Further collapsed - No last mile, mostly B2B circuits and often without NAT/QoE
  - **Peering ISP** – Further collapsed – Smaller last mile. Primarily B2B circuits, typically without NAT/QoE.
  - **Last mile ISP** – Further separated, lots of middle boxes, shaping, NAT and geographic dispersion lends well to separation
- Collapse when practical, separate as needed.
  - Some BNGs offer QoE, CGNAT and Agg/PE functions – this can become one layer but typically at a higher cost.



S2 | Design Examples

Practical examples for different types of ISPs

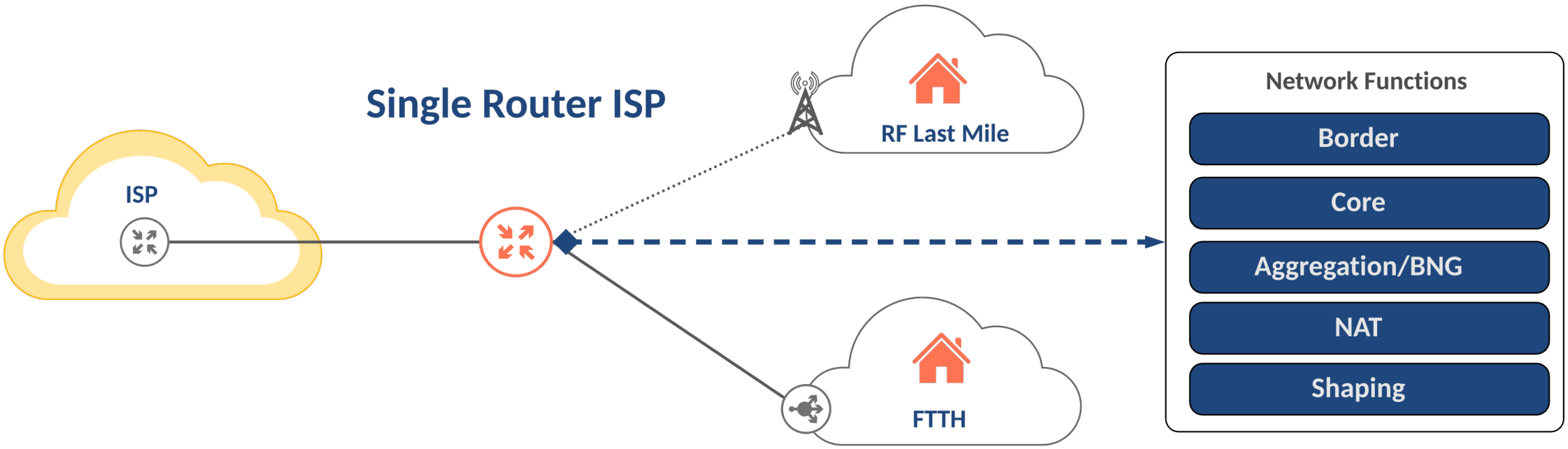
Section 2 | Design examples

02

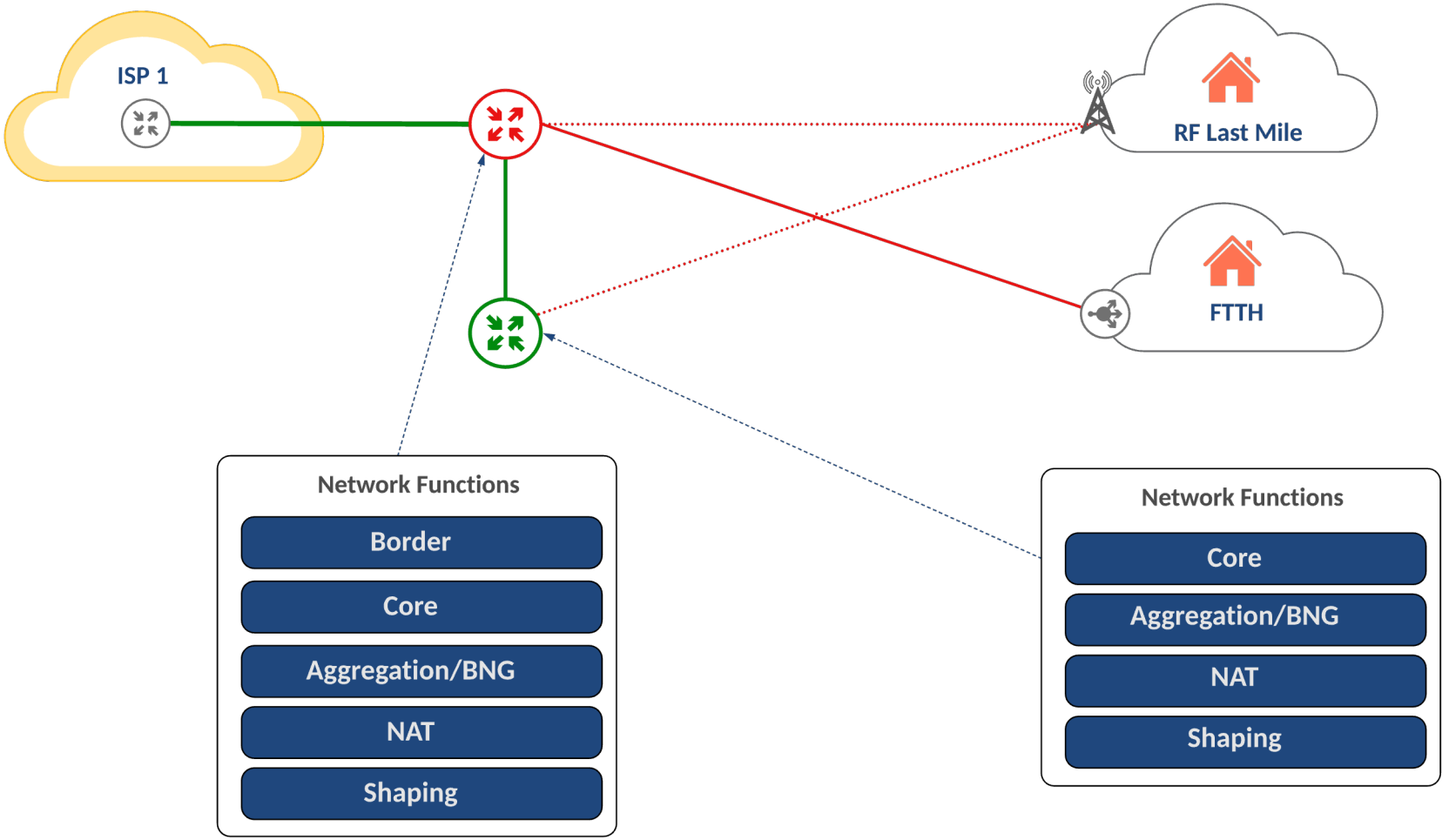
- Startup WISP or FISP
- Established WISP or FISP
- Telco
- Multiple Data Centers

Startup WISP or FISP

Keep it simple? - When first building a startup ISP, keeping it simple seems reasonable. As the ISP begins to grow, challenges with scale and design begin to arise. The network doesn't need to be needlessly complicated but should be built to the buisiness and technical requirements.



Scenario: Add a router to a network that's low on ports with high CPU



Issues with this design

Capacity ! - The main router is out of ports and is at 70% average CPU due to NAT and FW rules. Even if we solve the port issue, the problem of high cpu remains. Adding another router without adding another upstream doesn't make things better

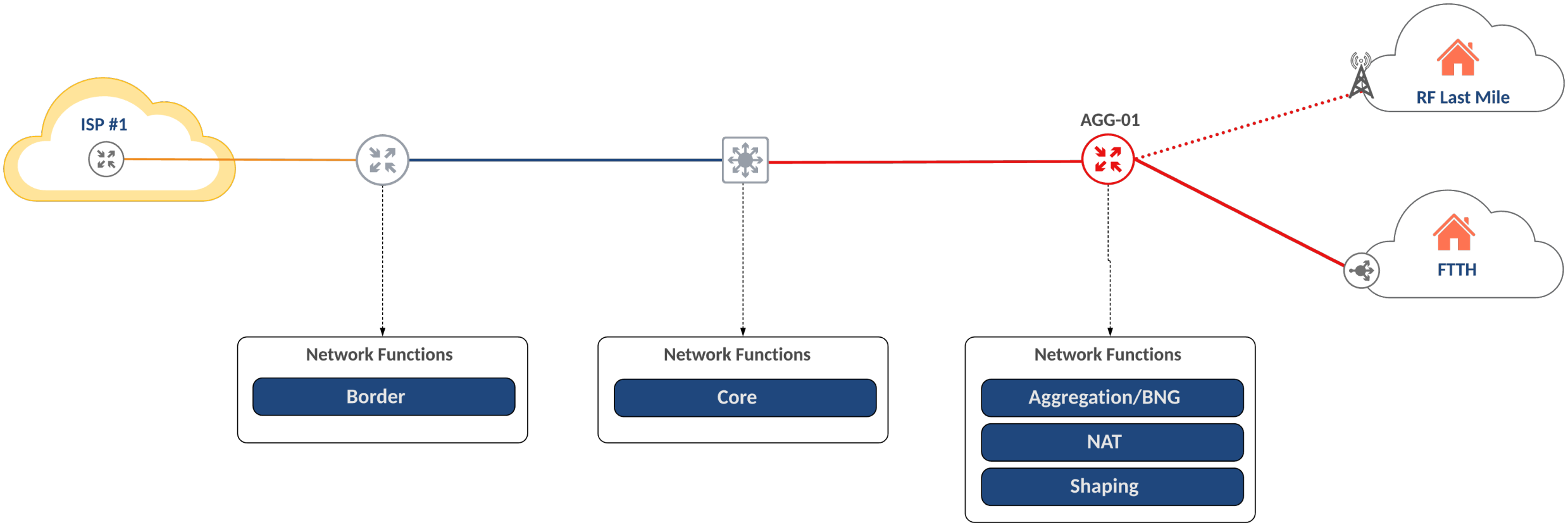
Scale ! - With low port availability and an overloaded router, this network cannot scale without adding a new upstream or a redesign.

Redundancy ! - Now there is a one-off design that creates a single point of failure.

Complexity ! - Functions are now inconsistent between the routers which creates complexity that will affect stability and raise opex.

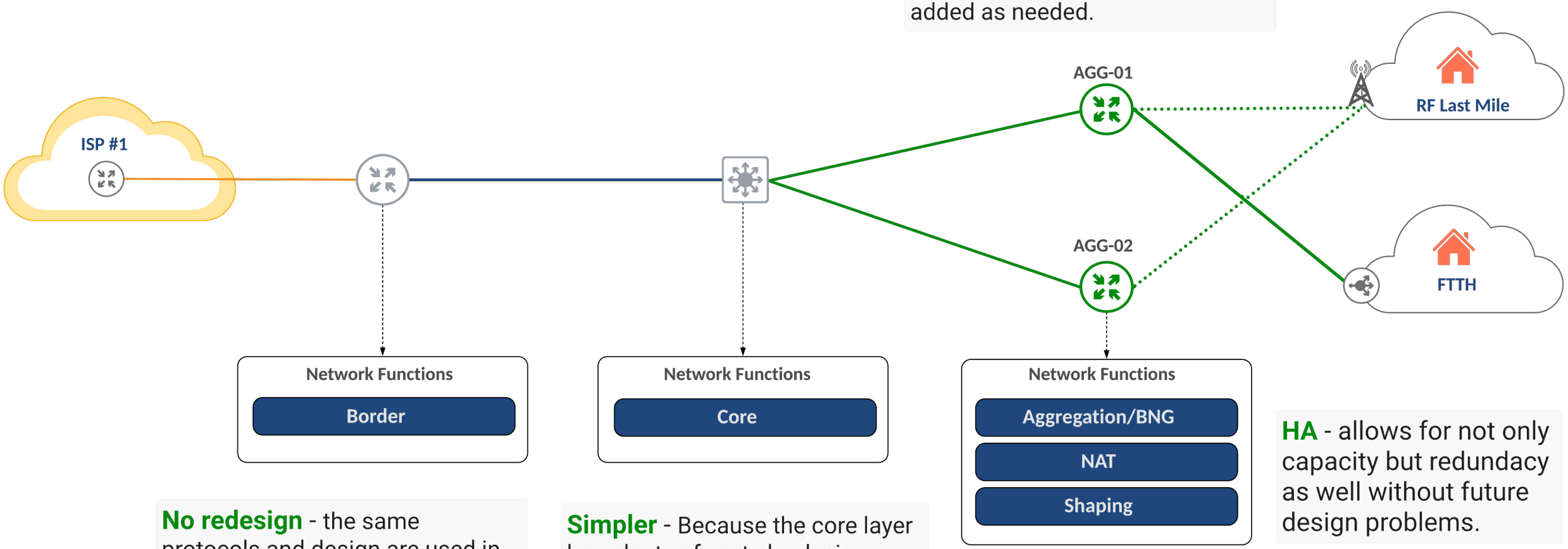
Same scenario in a network with distributed functions:

The AGG-01 router has an overloaded CPU and is low on ports



Solution: Add a second AGG router

Scalable - More agg pairs can be added as needed.



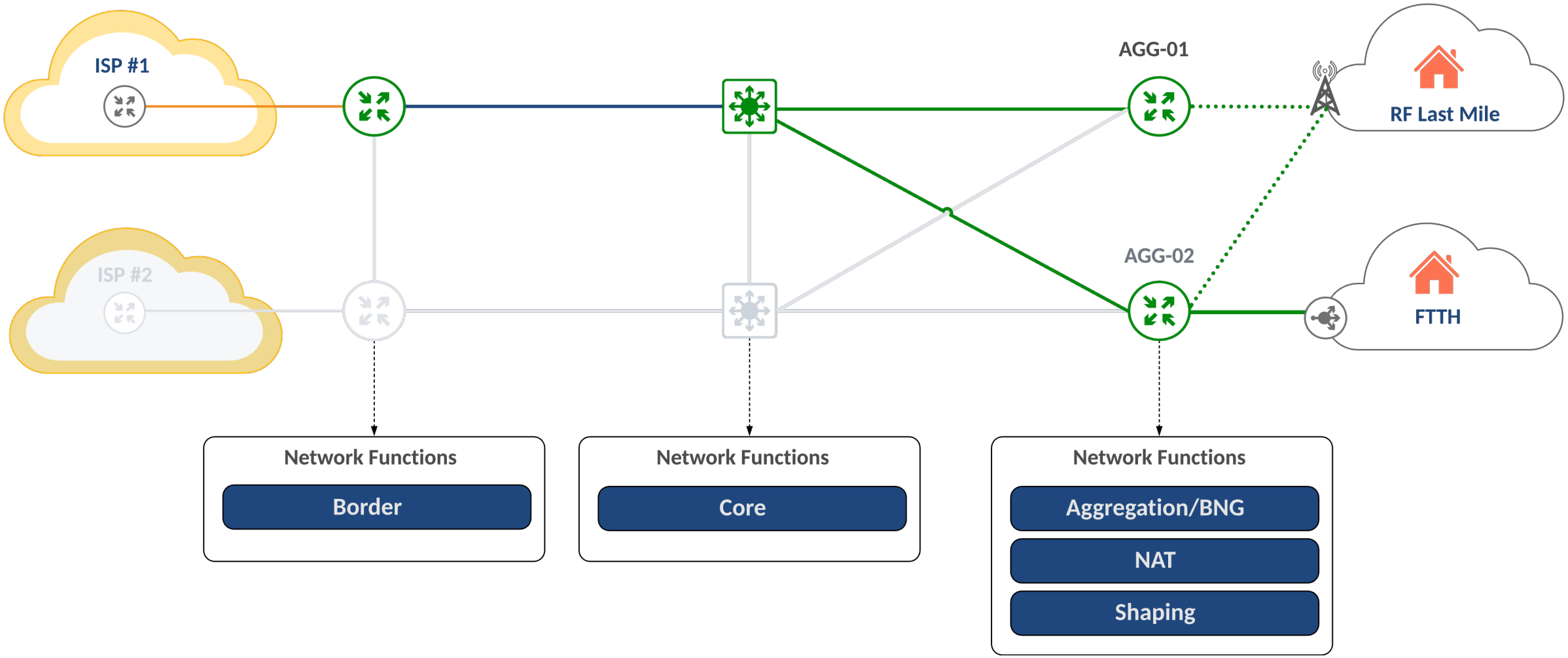
HA - allows for not only capacity but redundancy as well without future design problems.

No redesign - the same protocols and design are used in the new agg router. The disruption to the network is much lower.

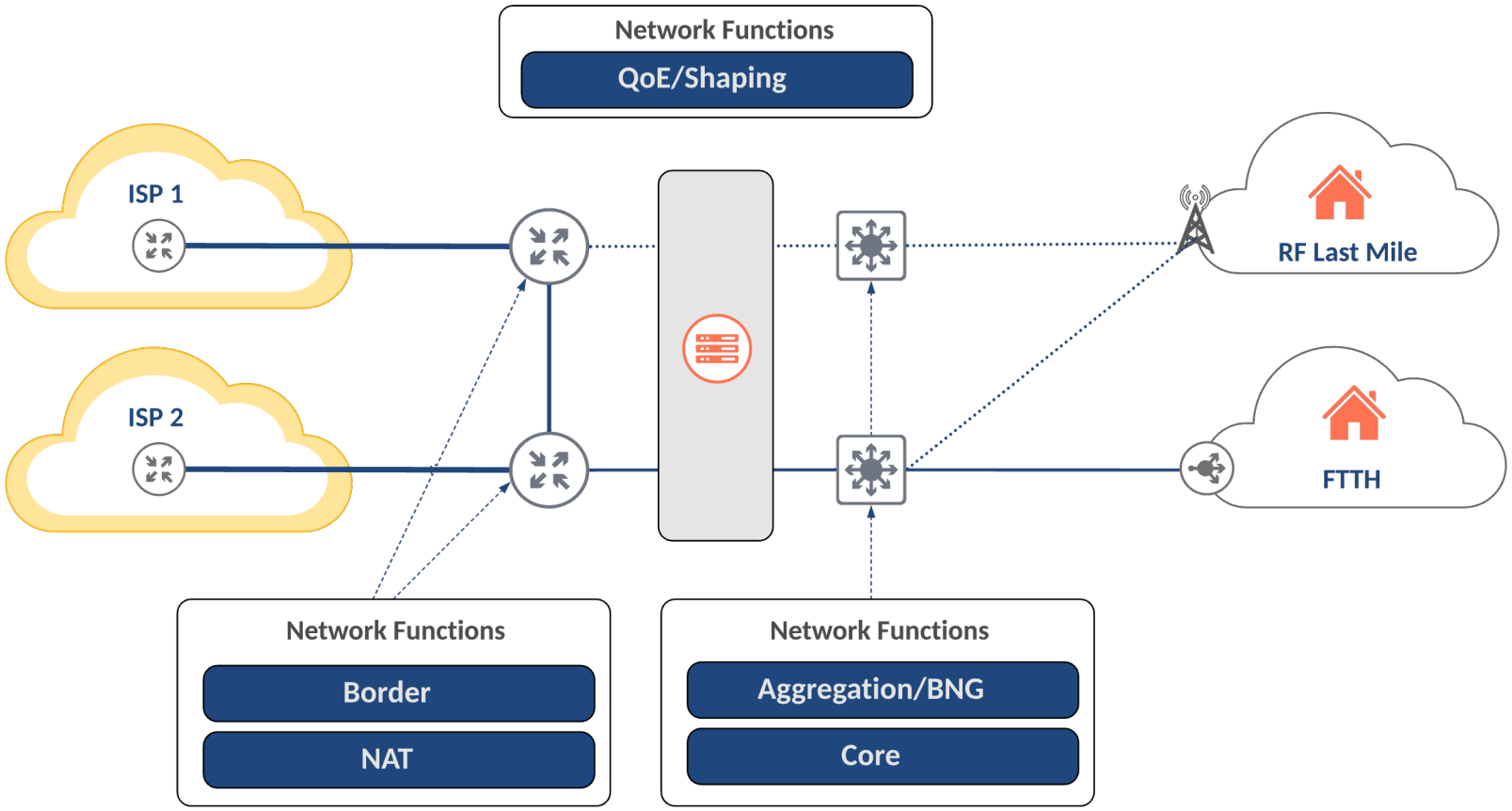
Simpler - Because the core layer has plenty of ports by design, additions are not hard.

Startup WISP or FISP - future planning

Highly available that's affordable - The biggest challenge with a redundant core network is often figuring out how to pay for it. By leveraging this design, not only can we solve the agg problem, we can also prepare for future additions without redesign



Scenario: Add QoE to an established WISP or FISP



Issues with this design

NAT ! - NAT is always the biggest challenge when adding QoE because the shaping appliance needs to see the original source of traffic. In this network, the core switches aren't ideal for NAT so the borders will need to perform NAT

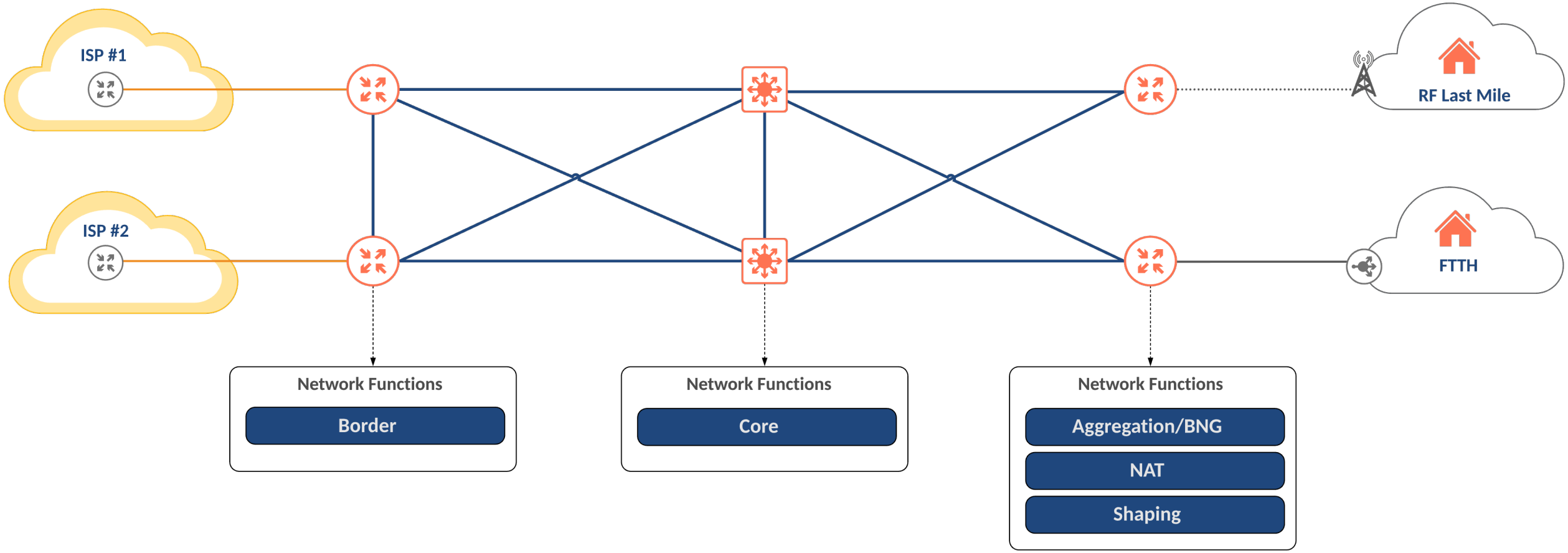
State ! - Ideally border routers shouldn't maintain state tables but these will have to becuase of NAT

Redundancy ! - Failover is now harder because state is involved and symmetric routing is needed for the borders

Scale ! - Adding another border router becomes challenging due to NAT/State issues.

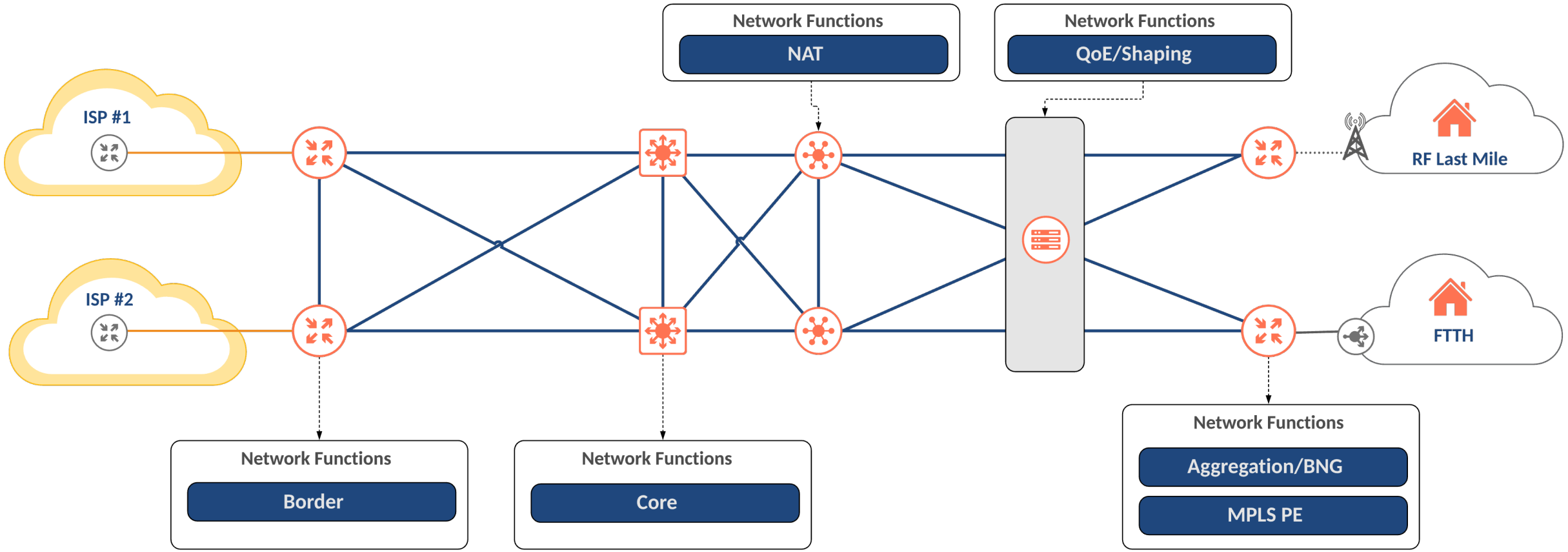
MPLS ! - There is no ideal place for VPLS termination from the last mile in this design.

Same scenario in a network with separated functions:
Add a QoE shaping appliance



Solution: add QoE/NAT layers

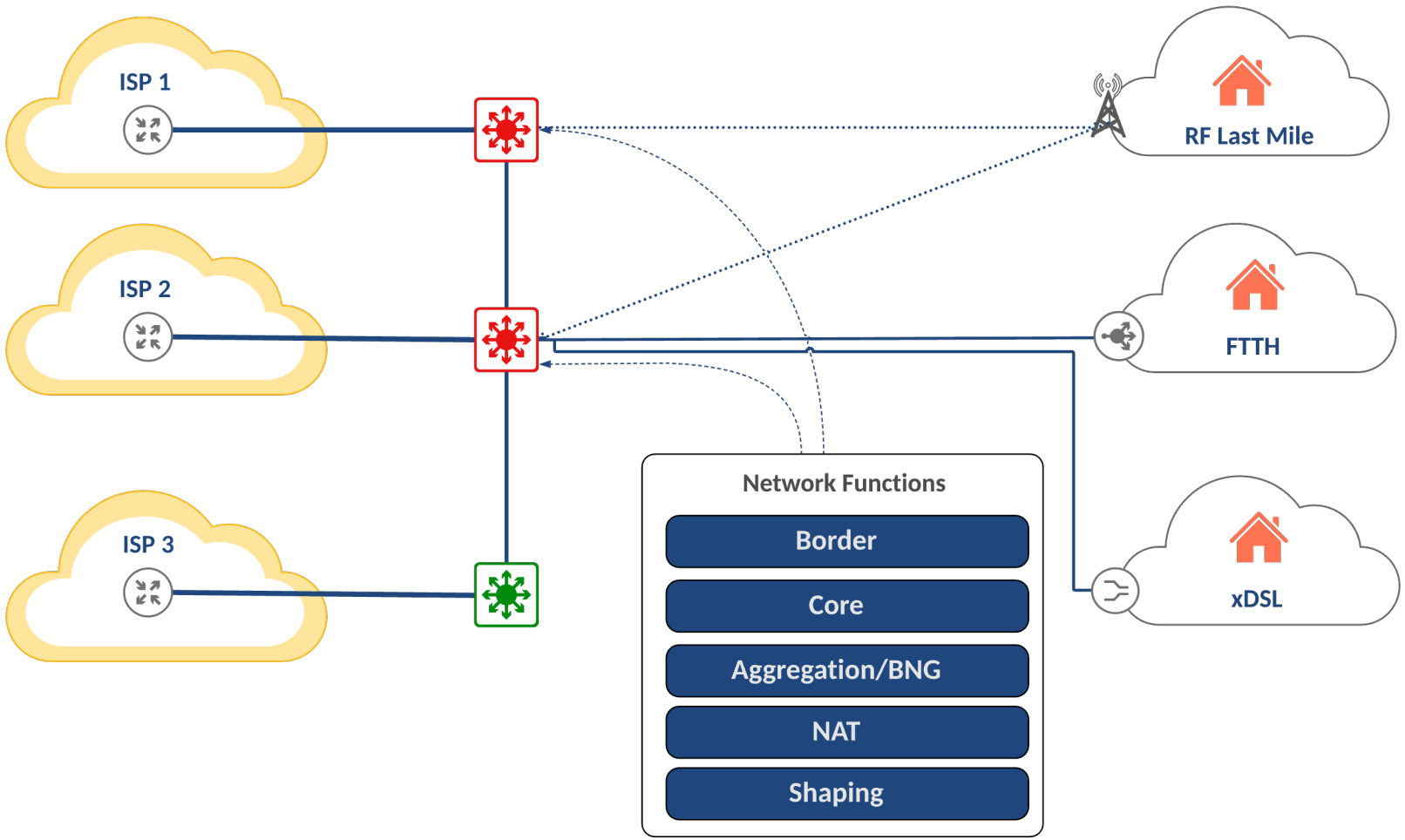
Scalable - More QoE shapers/CGNAT gateways can be added as needed to grow with agg pairs



No state issues at the border - Transit/IX circuits can go in/out the best path via routing without the overhead of maintaining state and symmetric routing

VPLS termination - VPLS can easily be implemented before the QoE shaper

Scenario: Add upstream capacity & table space for a telco



Issues with this design

Table space - L3 chassis switch 1 and 2 are running out of table space and a 3rd L3 chassis with more FIB space is added to peer with a new upstream. Peerings between the switches must filter prefixes to avoid FIB exhastion

State ! - Ideally border routers shouldn't maintain state tables but these will have to becuase of NAT

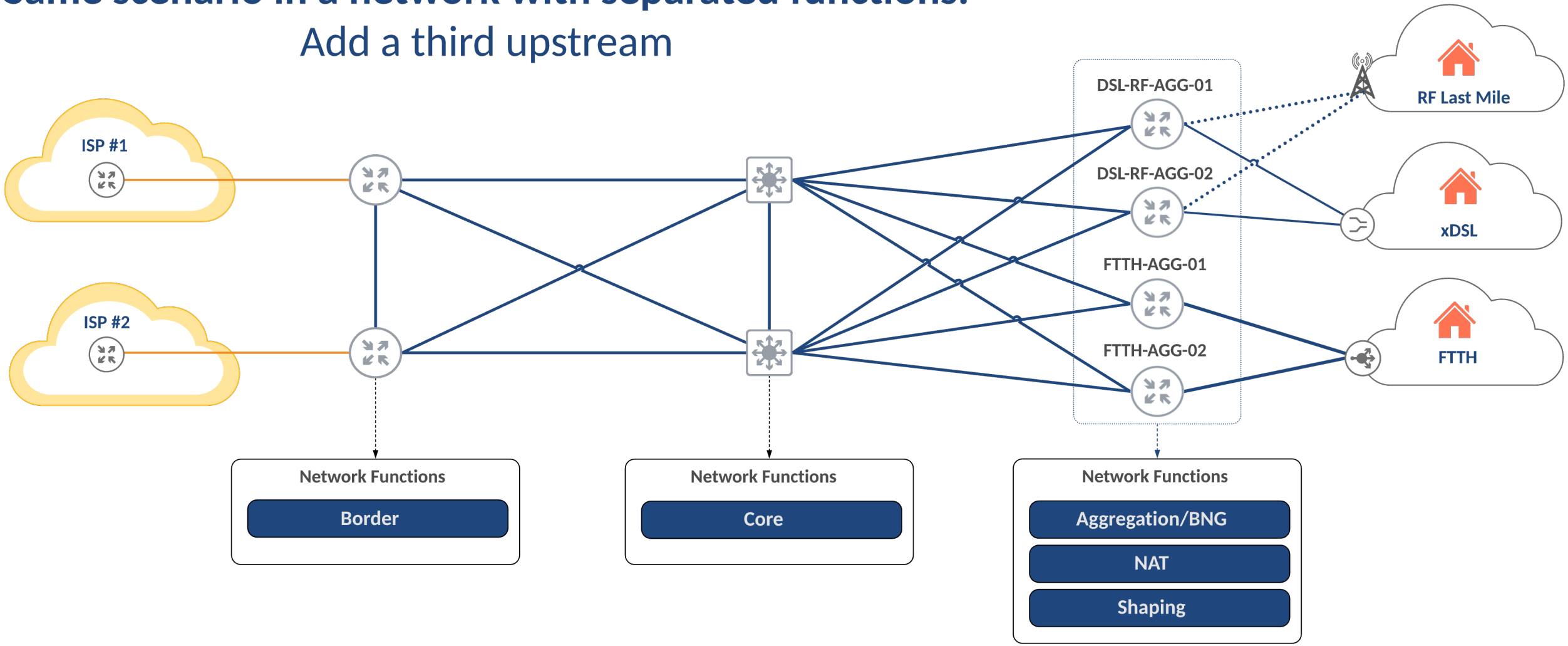
Last mile ! - Last mile circuits now have to be rebalanced across all 3 switches

Scale ! - Adding another border router becomes challenging due to table size, last mile connections and state

Cost ! - Adding L3 switches capable of full tables is expensive and the design still doesn't grow well

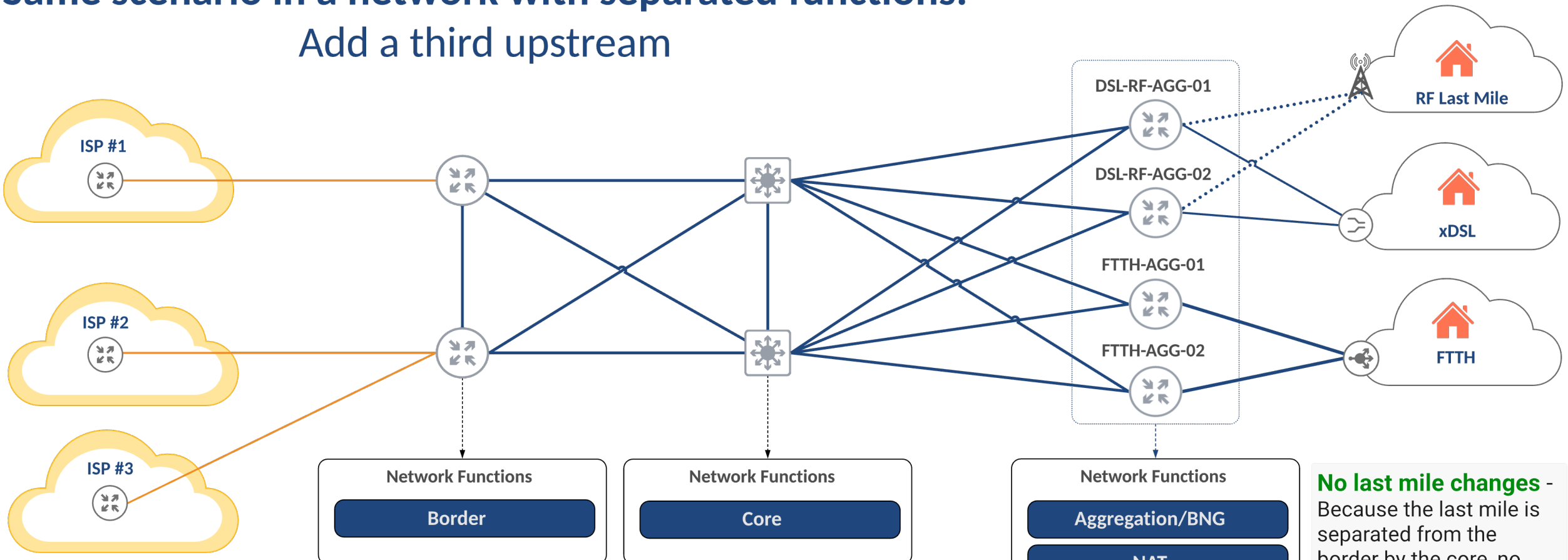
Same scenario in a network with separated functions:

Add a third upstream



Same scenario in a network with separated functions:

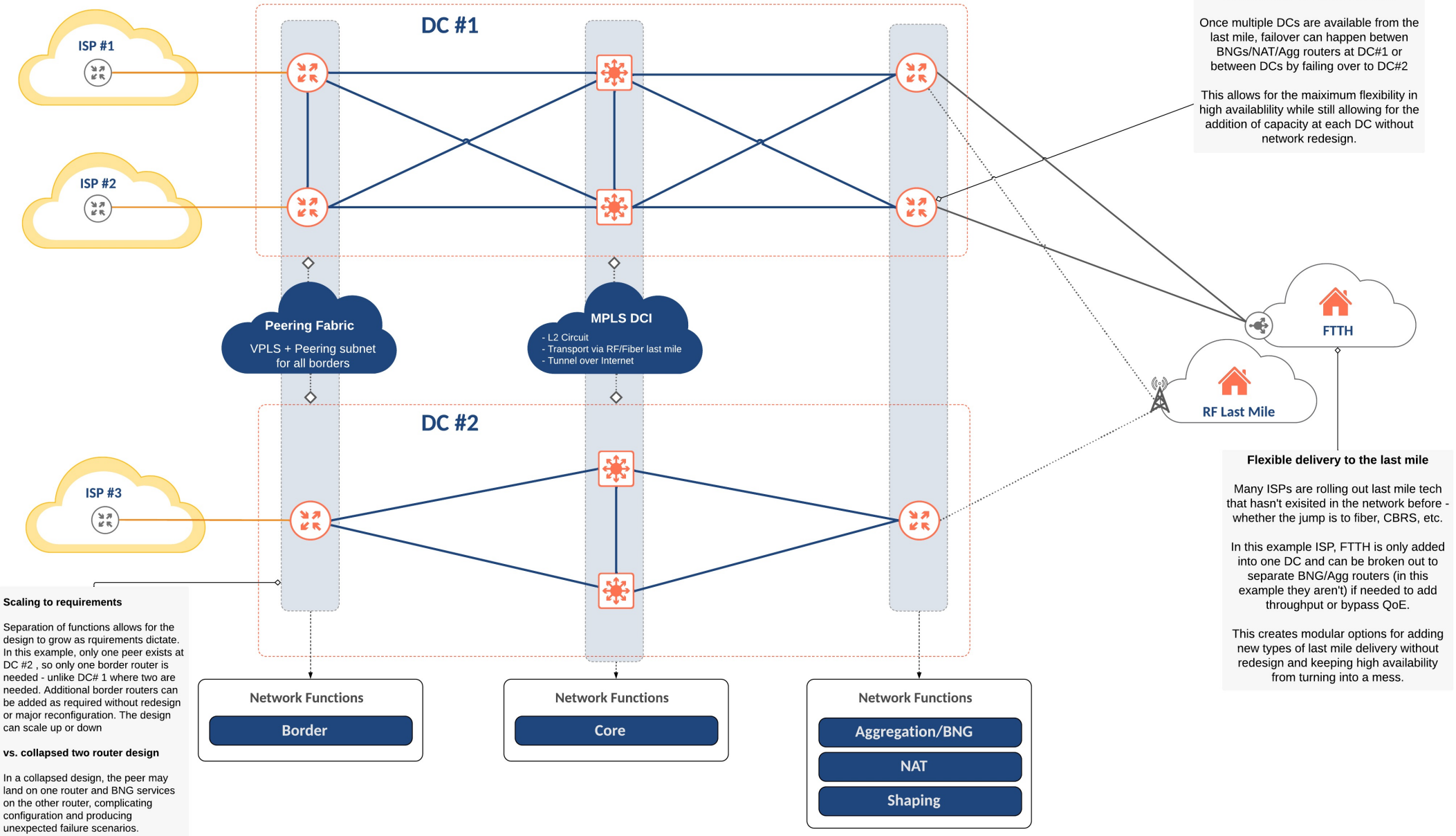
Add a third upstream



Use smaller border routers with more FIB - instead of larger expensive L3 switch chassis, this design uses 1U border routers and can add a 3rd circuit without issue

No last mile changes - Because the last mile is separated from the border by the core, no changes are needed even if a router is added.

Multiple DCs - Highly Available ISP with separated network functions



BNG/NAT/Agg Failover within the DC or between DCs

Once multiple DCs are available from the last mile, failover can happen betwen BNGs/NAT/Agg routers at DC#1 or between DCs by failing over to DC#2

This allows for the maiximum flexibility in high availablility while still allowing for the addition of capacity at each DC without network redesign.

Flexible delivery to the last mile

Many ISPs are rolling out last mile tech that hasn't exisited in the network before - whether the jump is to fiber, CBRS, etc.

In this example ISP, FTTH is only added into one DC and can be broken out to separate BNG/Agg routers (in this example they aren't) if needed to add throughput or bypass QoE.

This creates modular options for adding new types of last mile delivery without redesign and keeping high availability from turning into a mess.

Scaling to requirements

Separation of functions allows for the design to grow as rquirements dictate. In this example, only one peer exists at DC #2 , so only one border router is needed - unlike DC# 1 where two are needed. Additional border routers can be added as required without redesign or major reconfiguration. The design can scale up or down

vs. collapsed two router design

In a collapsed design, the peer may land on one router and BNG services on the other router, complicating configuration and producing unexpected failure scenarios.

S3 | Equipment

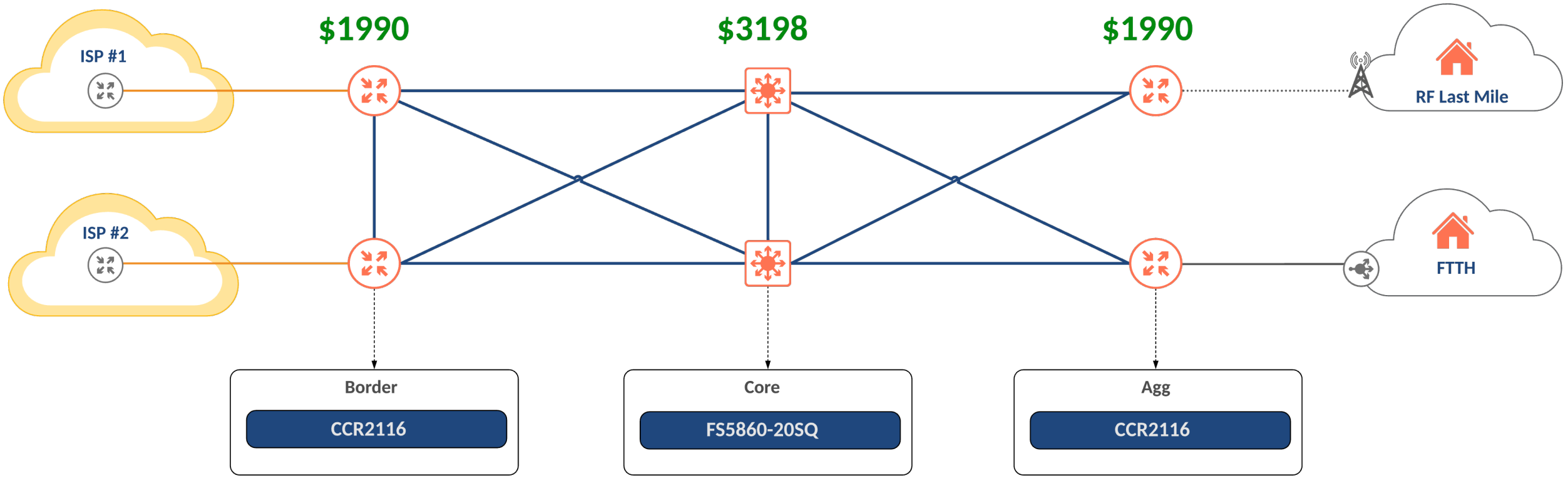
Overview of commodity equipment and costs

Section 3 | Equipment and budget

03

- Cost exercise for 80G ISP
- MikroTik
- Fiberstore
- IP Infusion + Edge Core / Ufispace

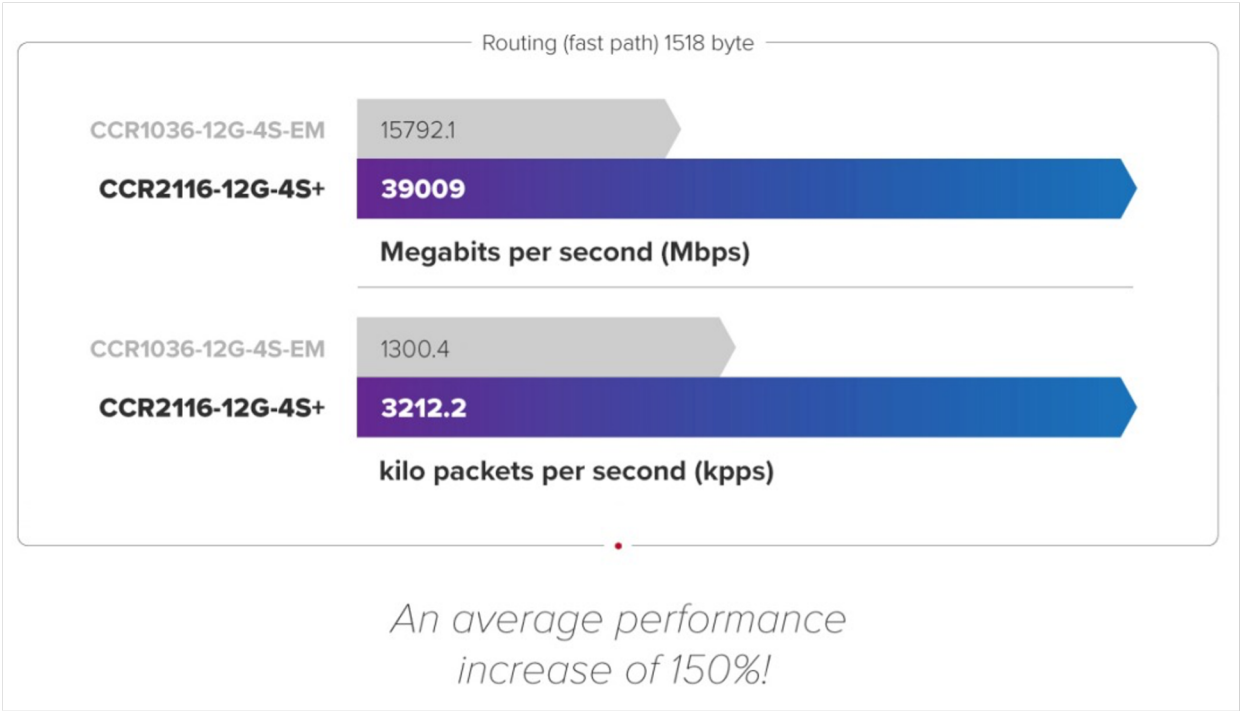
CAPEX for 80G of throughput $7178



Discussion: What are my options for 10G/100G routers ?

- The state of 10/100G routers for WISP/FISP and Telco operators
  - **MikroTik** – just released the first of their 100G equipment line. Plenty of 10G options
  - **Ubiquiti** – no current 100G solution and the routing stack is years behind MikroTik. One 10G router exists.
  - **X86** – very practical and affordable option – can use CHR, FRRouting, VyOS, DanOS
  - **Whitebox** – No true "router" exists with a CPU but L3 switches with extended TCAM can be used
  - **Juniper/Arista/Nokia** – Only a handful of choices under $25k - $50k and lead times are 1 to 2 years long due to the chip shortage.
  - **! Be super careful !** about eBAY & graymarket as compromised/counterfeit network devices are now being sold so they can be remotely accessed by hackers for attacks.

Routing - MikroTik CCR2116



Overview - A new 4 x10G, 12 x 1G router from MikroTik that uses a 16 core Amazon ARM64 CPU capable of 60G+ and a Marvell ASIC capable of 100G

Considerations

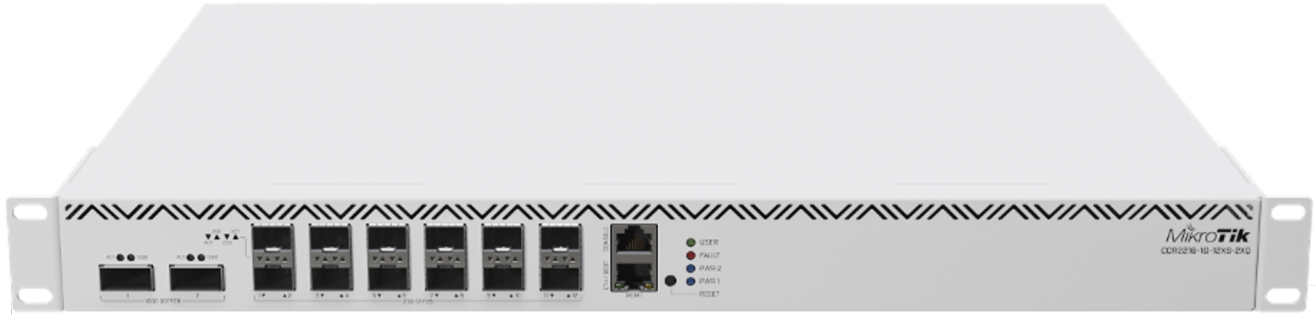
- Positioned as a CCR1036 replacement. A great successor with the addition of an ASIC and a price tag under $1000

- ROSv7 should be stable by June/July and it fills in a lot of protocol gaps like iBGP in IPv6 and can now pass RFC2544 (and newer) testing.

- Initial testing on BGP full tables is actually faster than Cisco/Juniper/Arista.

Routing - MikroTIk CCR2216

Overview - A new 2 x100G, 12 x 25G router from MikroTik that uses a 16 core Amazon ARM64 CPU capable of 200G and an Marvell ASIC capable of 600G



CCR2216-1G-12XS-2XQ

1518 byte

Mode	Configuration	kpps	Mbps
Bridging	Switching	**16254.8+**	**197398.3+**
Routing	L3HW	**16254.8+**	**197398.3+**
Routing	25 ip filter rules + L3HW	**16254.8+**	**197398.3+**

**Tested with the Xena 200Gbps test platform*

With L3HW offloaded connections, this CPU can be at least 4x faster than the CCR1072 Tile CPU!

Considerations

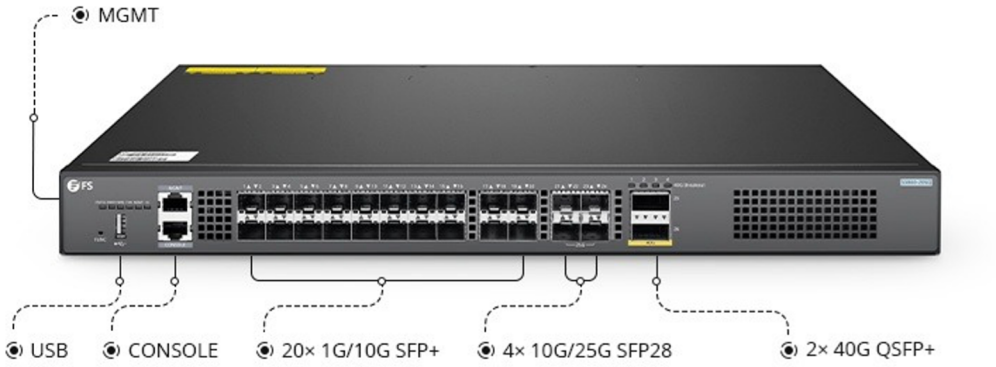
- This is a new class of router - nothing like this exists in other vedors with a multi 100G CPU and an ASIC. Creates massive flexibility

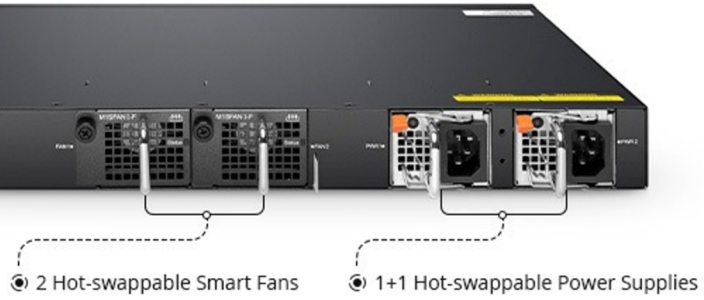
- ROSv7 should be stable by June/July and it fills in a lot of protocol gaps like iBGP in IPv6 and can now pass RFC2544 (and newer) testing.

- Initial testing on BGP full tables is actually faster than Cisco/Juniper/Arista.

Discussion: What are my options for 10G/100G L3 switches ?

- The state of 10G/100G L3 switches for WISP operators
  - **MikroTik** – 100G CRS 5xx 4 port 100G – 8 to 16 port count expected
  - **Ubiquiti** – no current 10/100g solution for L3 switching
  - **FS** – Multiple options for (10G x 48, 8 x 100G) and (32 x 100G)
  - **Whitebox** – Multiple options for (10G x 48, 6 x 100G) and (32 x 100G)
  - **Juniper/Arista/Nokia** – Only a handful of choices under $25k - $50k and lead times are 1 to 2 years long due to the chip shortage.
  - **! Be super careful !** about eBAY & graymarket as compromised devices are now being sold.





L3 switching - FS 5860-20SQ

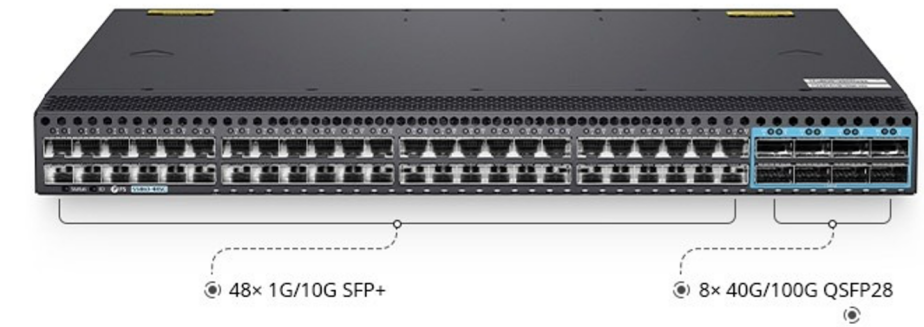
Overview - A 20 x 10G, 4 x 25G, 2 x 40G L3 switch that is capable of BGP, OSFP, IS-IS for IPv4 and IPv6. Supports both stacking and MLAG. Uses IPI ZebOS under the hood.

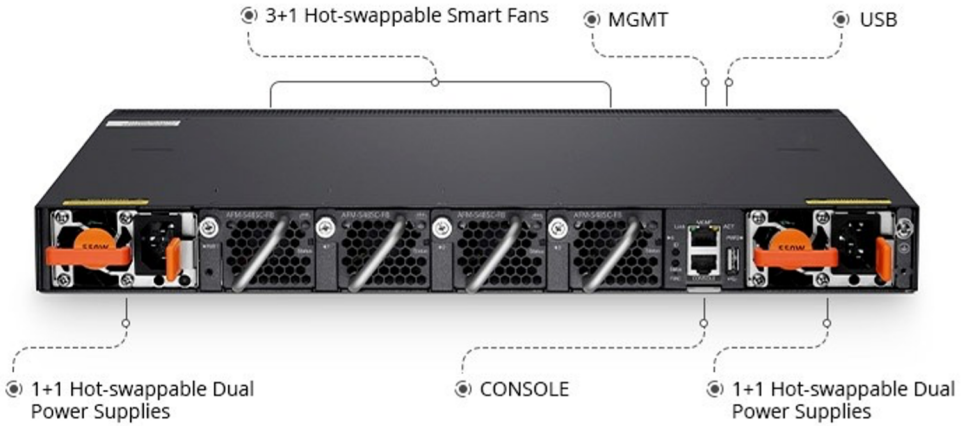
Considerations

- This is a fantastic and affordable ($1500) core switch that can be used in routed, SWC or Hybrid topologies

- Great IPv6 support not just for forwarding but also for management.

- Have deployed this in prod using SWC design for a 3,000 sub WISP/FISP with 10Gbps of traffic.





L3 switching - FS 5860-48SC

Overview - An 8 x 100G, 48 x 10G L3 switch that is capable of BGP, OSFP, IS-IS for IPv4 and IPv6. Supports both stacking and MLAG. Uses IPI ZebOS under the hood.

Considerations

- This is a fantastic and affordable ($4k) core switch that can be used in routed, SWC or Hybrid topologies

- Great IPv6 support not just for forwarding but also for management.

- Have deployed this in prod using SWC design for a 5,000 sub WISP/FISP with 20Gbps of traffic.

OCP ACCEPTED

L3 switching - Edge Core 5912-54X

Overview - A 48 x 10G, 6 x 100G L3 switch that is capable of BGP, OSFP, IS-IS for IPv4/IPv6. MPLS capable with Segment Routing. Uses IPI OcNOS-SP for NOS software.

Considerations

- This is the most common 10G/100G L3 switch we deploy for WISPs

- Can be used in a routed or hybrid topology.

- OcNOS-SP from IP Infusion is a fantastic solution when higher end protocols like Segment Routing are needed but Juniper/Arista/Nokia are too expensive or unavailable.

S9600-32X

32-Port, 25/100GE Open Aggregation Router

L3 switching - UfiSpace 9600-32X

Overview - A 32 x 100GL3 switch that is capable of BGP, OSFP, IS-IS for IPv4/IPv6. MPLS capable with Segment Routing. Uses IPI OcNOS-SP for NOS software.

Considerations

- This is one of the most affordable 100G core switches that supports SR-MPLS

- Can be used in a routed or hybrid topology.

- OcNOS-SP from IP Infusion is a fantastic solution when higher end protocols like Segment Routing are needed but Juniper/Arista/Nokia are too expensive or unavailable.

Thank you!

Thank you for joining us today! This is a large topic with plenty of nuances, if you'd like to brainstorm with us your deployment, network architecture, or software ecosystem, do not hesitate to contact us using the information below.

We are a full-service networking firm that can help identify areas of improvement, design network architecture, as well as plan and execute your migration windows.

IP ArchiTechs
consulting@iparchitechs.com
11757 W Ken Caryl Ave, Littleton, CO, 80127
+1 (855) 645-7684