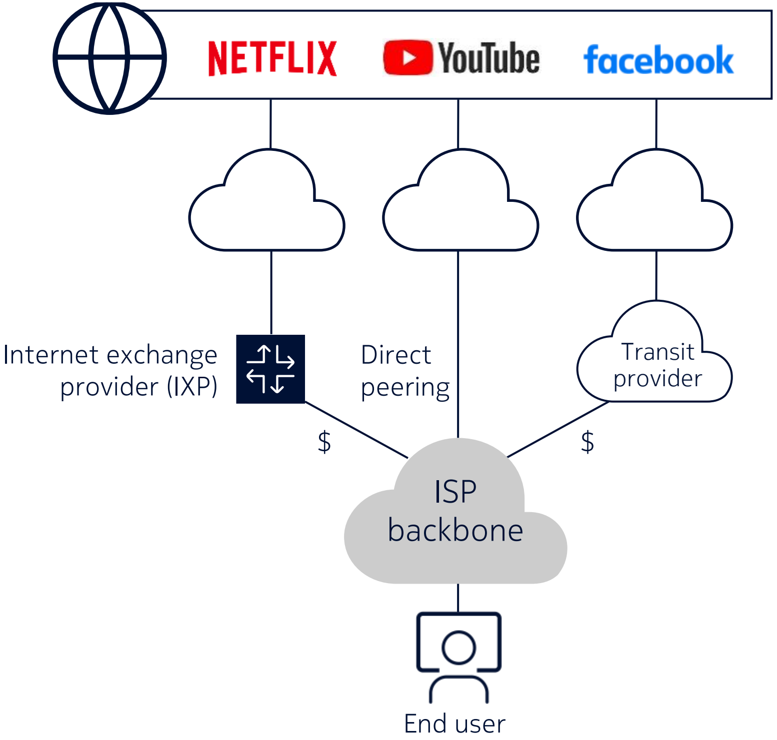
Evolve your IP network peering

Daniel Beszkid

10 October 2024

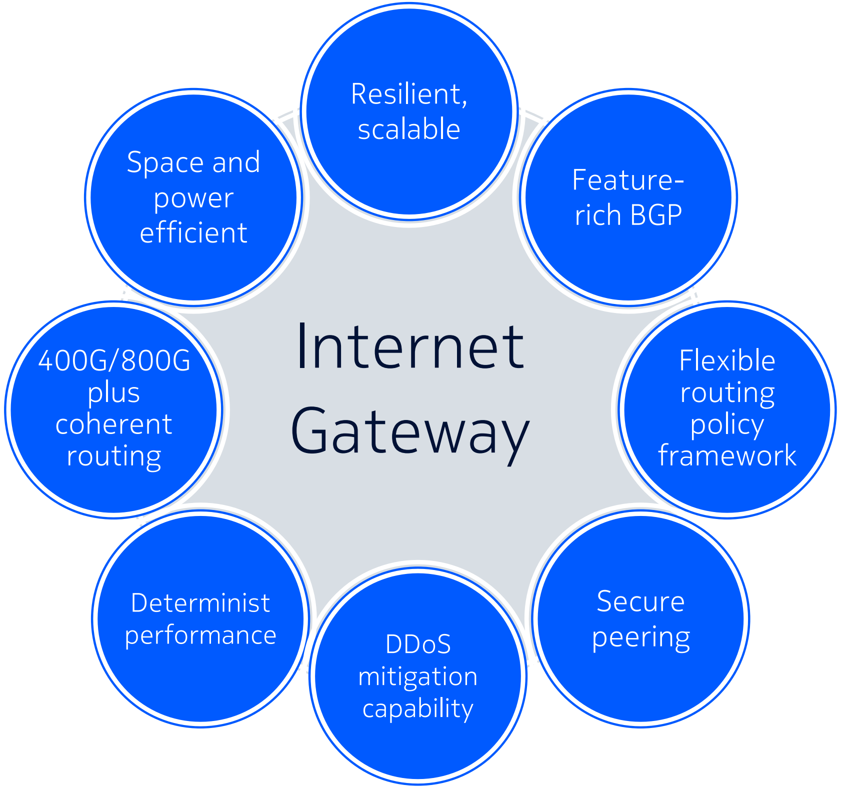
An introduction to peering

- Peering router, also known as internet gateway, primary responsibility is to facilitate the exchange of traffic between two or more autonomous networks (ASes)
- Peering to other networks (peers) can be via
  - Direct/private peering, typically without any monetary exchange
  - Internet exchange peering in a co-location space where network operators exchange traffic via IXP L2 network (paid service)
  - Transit providers (paid service)
- Routes exchange is done via BGP protocol
- Additionally, peering routers should also play a crucial role preventing DDOS attacks and BGP route hijacking



What would be a best in class peering router

1. Resilient, scalable, and feature-rich BGP process
2. Flexible routing policy framework for frequent route policy updates
3. Secure routing and DDoS mitigation on network entry
4. Guaranteed performance for high number of flows and filters
5. Space and power efficient design with
   - High speed universal interfaces
   - Coherent routing support



1. BGP capabilities

Robust and feature-rich

Scale
N x k BGP sessions
Multiple copies of Internet routing table (RIB)
Millions of routes in the FIB (IPv4/IPv6)

Error handling
Avoiding disruptive session resets using attribute-discard, treat-as-withdraw, AFI/SAFI disable.
Optional Graceful Restart Notification.

Path visibility and reconvergence
Peer & Next-Hop tracking, Edge-PIC, Next-Hop Indirection, ADD-PATH & Multipath, Rapid Update & Withdraw. Incremental Prefix-List Updates.

OS

High availability
Fully NSR capable.
No need for Graceful Restart, but can be GR helper.

Robust
Field-proven BGP implementation and software quality assurance

RIB-OUT processing
Avoid head-of-line blocking caused by slow peers using Update message per-peer

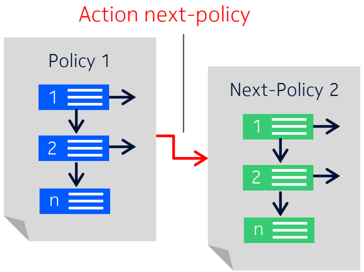
2. Route policy framework

Flexible and scalable route policy language

Base policy can be extended using advanced features to provide scalability and simplification

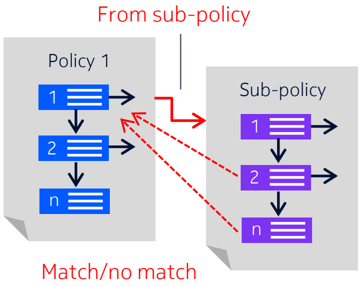
Nested policies

Allows for policy entries to reference a **next** policy on a match (policy chaining). Up to five nested policies can be configured.



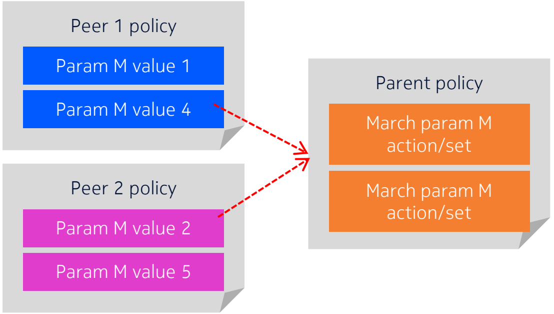
Scaling with sub-routines

Policy statement that calls another policy (**sub-policy**). Sub-policy allows for commonly used match/action criteria to be used by multiple policies.



Simplification with parameterization

Parent policy references parameters in each **child policy**, even though the value of those parameters is different for each policy.



“**Policy test**” allows for test evaluation of policy before applying

2. Route policy testing

Example

small handy feature allows to test and evaluate route policies before they are applied to BGP

```
A:admin@SR1# show router bgp policy-test plcy-or-long-expr "EXT-AS-IMPORT" family ipv4
prefix 0.0.0.0/0 longer neighbor 192.168.0.3
===============================================================================
 BGP Router ID:10.0.0.1         AS:64501       Local AS:64501
===============================================================================
 Legend -
 Status codes  : u - used, s - suppressed, h - history, d - decayed, * - valid
                 l - leaked, x - stale, > - best, b - backup, p - purge
 Origin codes  : i - IGP, e - EGP, ? - incomplete

===============================================================================
BGP IPv4 Routes
===============================================================================
      Network                                                LocalPref   MED
      Nexthop                                                Path-Id     Label
      As-Path
-------------------------------------------------------------------------------
Accepted by Policy EXT-AS-IMPORT Entry Routes-AS64503
      10.10.1.24/29                                          None        None
      192.168.0.3                                            None        n/a
      64503                                                              -
Accepted by Policy EXT-AS-IMPORT Entry Routes-AS64503
      10.10.20.103/32                                        None        None
      192.168.0.3                                            None        n/a
      64503                                                              -
Accepted by Policy EXT-AS-IMPORT Entry Routes-AS64503
      192.168.0.0/24                                         None        None
      192.168.0.3                                            None        n/a
      64503                                                              -
-------------------------------------------------------------------------------
Routes : 3
===============================================================================
```

3. Secure peering to the hostile internet

Security pillars

DDoS

- Block DDoS at the network edge

Filters

- Control Plane filters for deny/permit access, queues for rate-limited access to the CPU
- Management Access filters

ACL scale

- Up to N x 10k filter policies per system, with up to N x 100k filter policy entries

Platform Trust

- Secure boot
- Router hardening guide and security BCP recommendations

Peering router security features

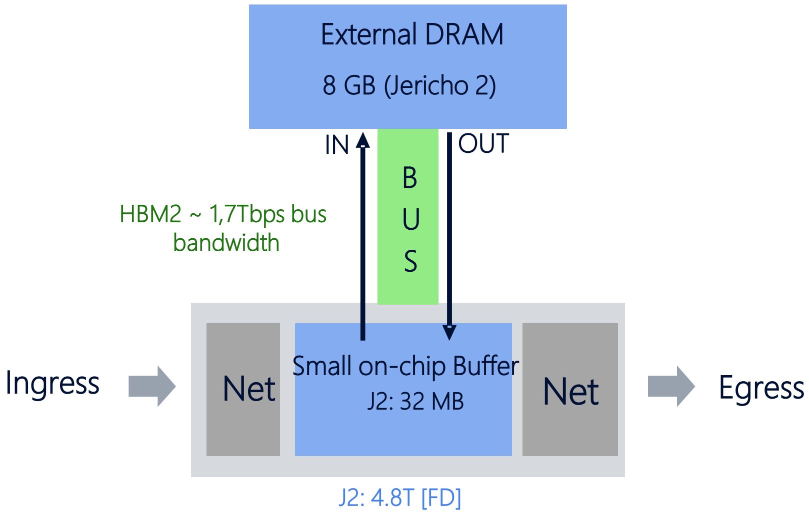
- RTBH, MD5 auth, GTSM, BGP FlowSpec (IPv4/IPv6), BGP Origin validation/RPKI etc.

Encryption

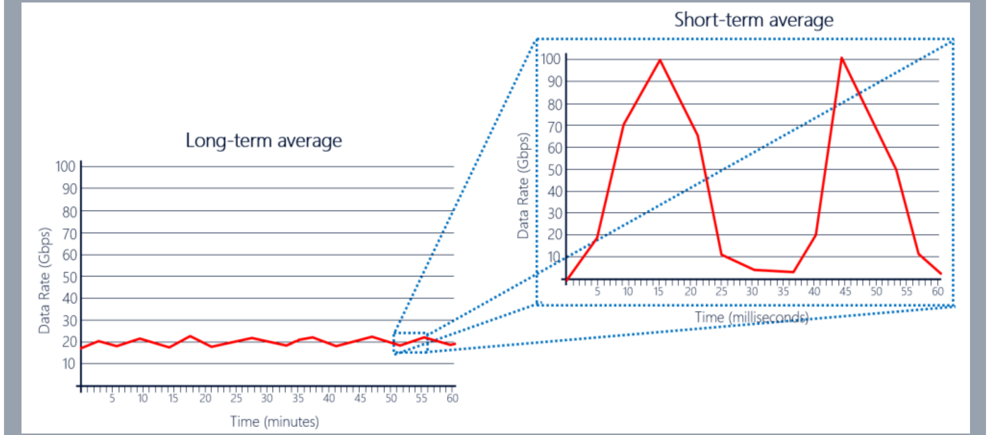
- MACSeC
- Quantum safe transport

4. Guaranteed performance for high number of flows and filters

Overcoming the ASIC-speed to memory-speed divergence



By definition, a microburst is a "short spike in network traffic", or "a burst of frames too short to be measured by standard polling mechanisms".

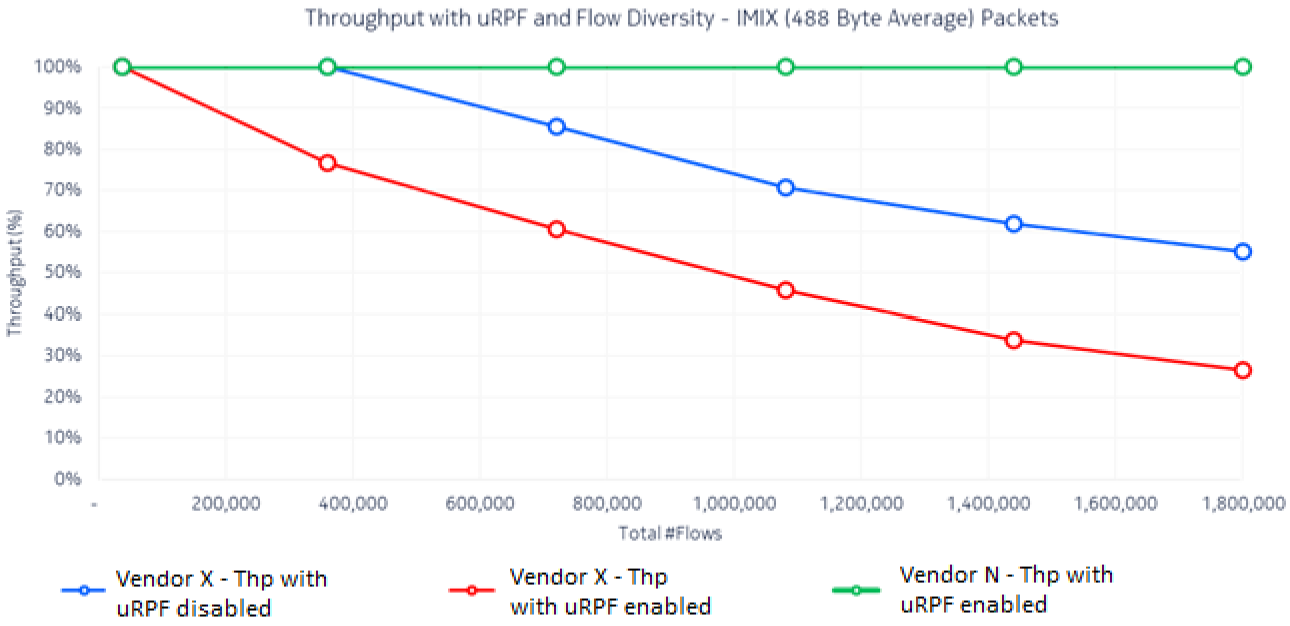


Bus bottleneck causing packet loss on the memory bus independently of any QoS under microbursts (traffic spikes in 1/1000th of seconds)

https://research.fb.com/wp-content/uploads/2017/10/imc17-final60.pdf

4. Guaranteed performance for high number of flows and filters

Performance with scale

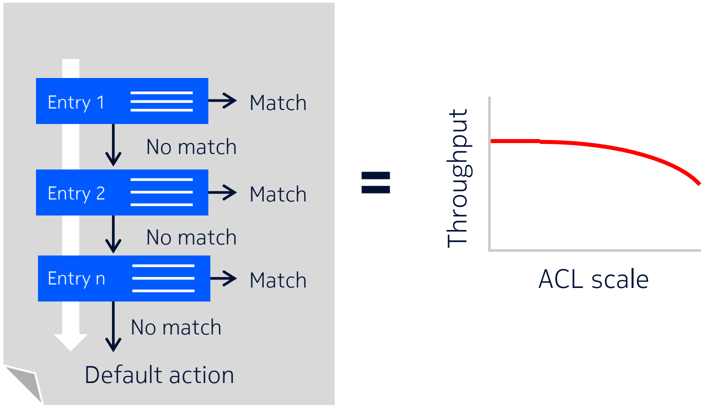


4. Guaranteed performance for high number of flows and filters

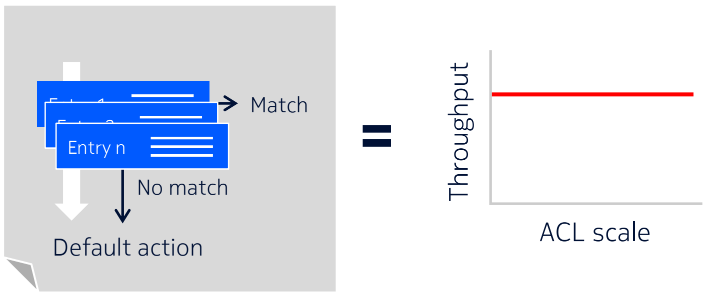
Performance with scale

The ability to apply large ACLs at scale is paramount for a peering router

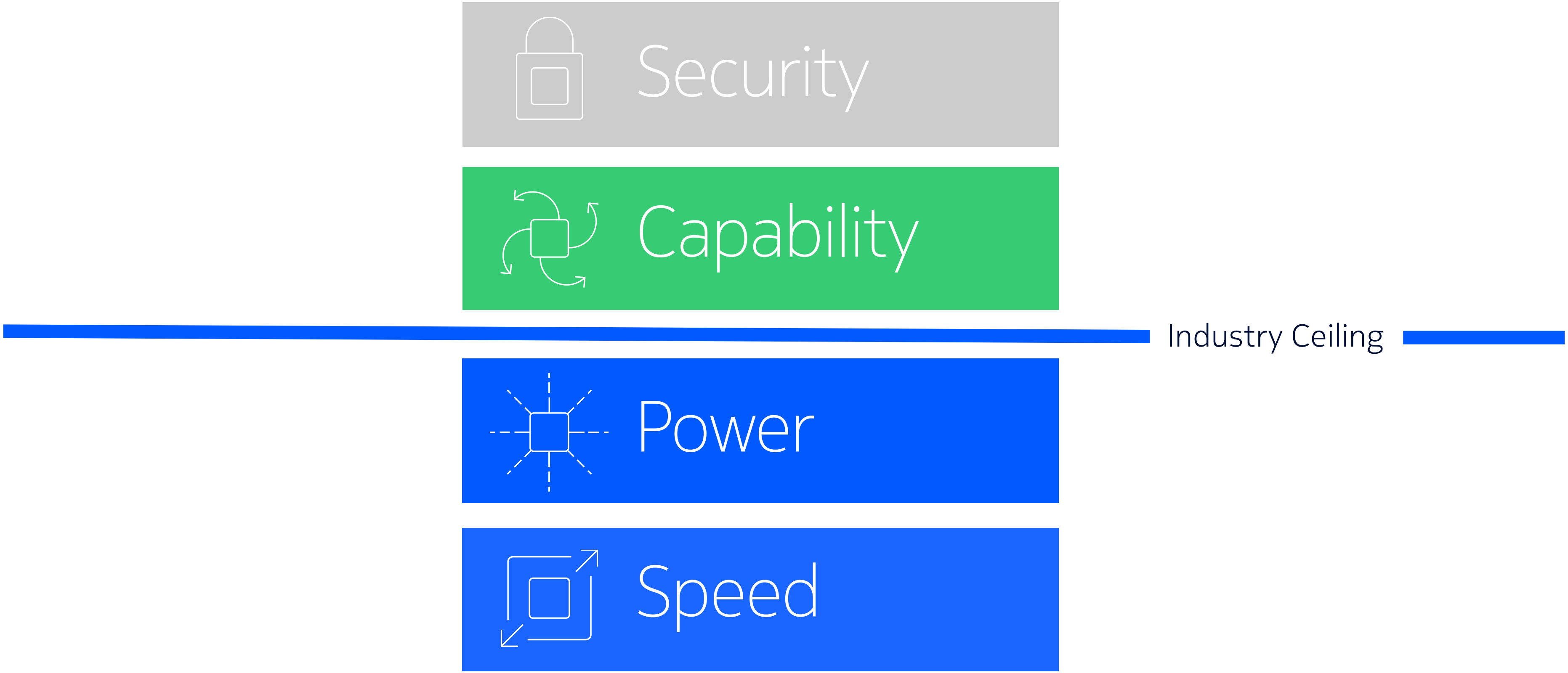
Parse through each entry sequentially may result performance degradation



ACL packet match in a single pass regardless of the number of ACL entries or ordering



5. Space and power efficient design



5. Space and power efficient design

High speed interfaces

In today modern routers the most efficient port cage utilization is key, thus universal port cages are must complemented with breakout support for investment protection

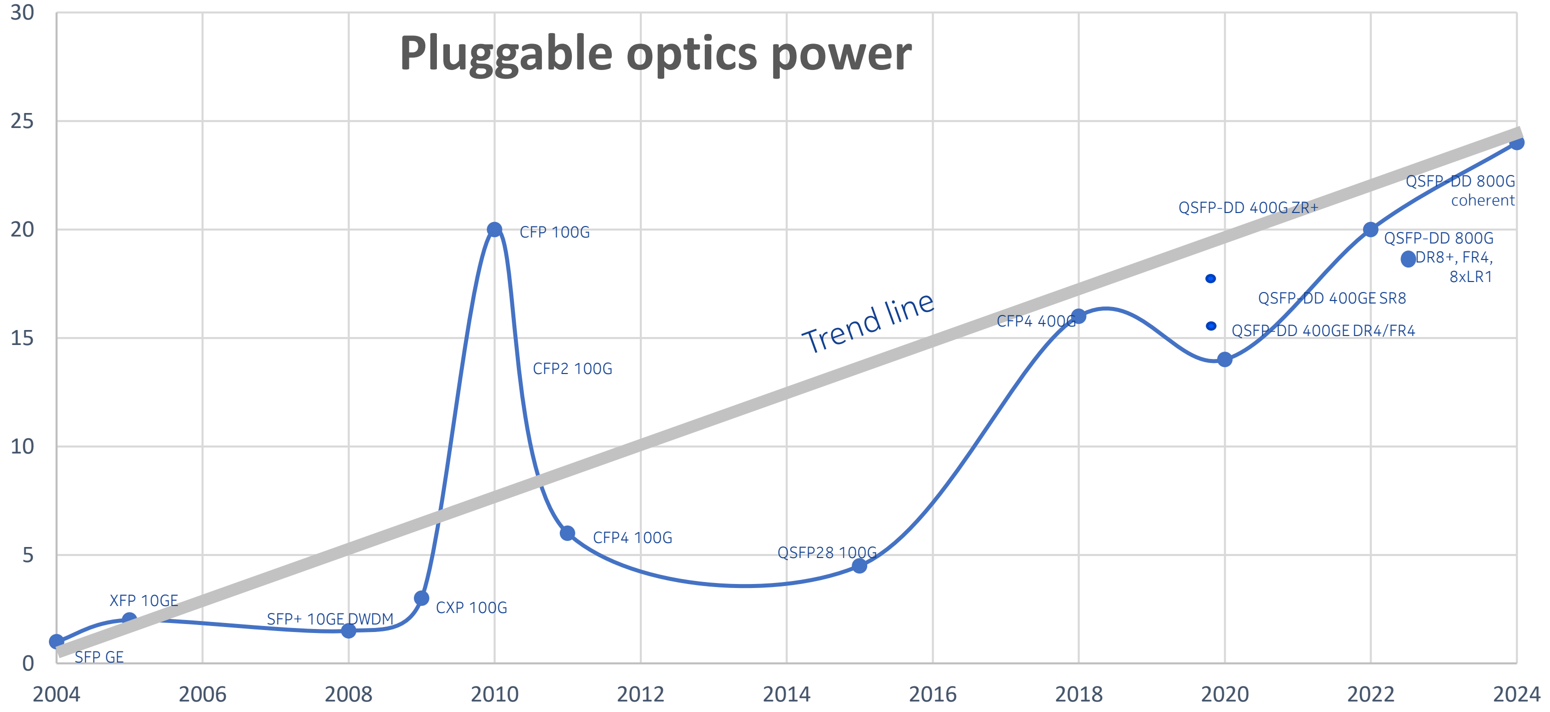
QSFP form factor: QSFP+, QSFP28, QSFP28-DD, QSFP56-DD, QSFP112, QSFP112-DD

SFP form factor: SFP+, SFP28, SFP56, SFP56-DD, SFP112, SFP112-DD

Breakout options: 4x 10G, 4x 25G, 10x 10G, 2x 100G, 4x 100G, 2x 400G, 8x 100G

5. Space and power efficient design

Power consumption of the optics

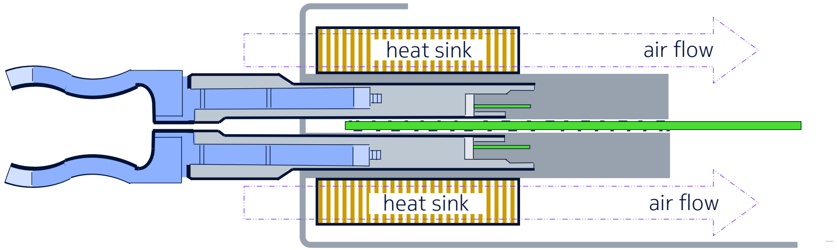


5. Space and power efficient design

Power consumption vs. cooling design

Cooling design has enormous impact to the power consumption:

- Belly-to-belly cages
- Dual-sided PCBs
- Honeycomb mesh air intakes



Belly-to-Belly cage design

5. Space and power efficient design

Power consumption calculation method accross vendors

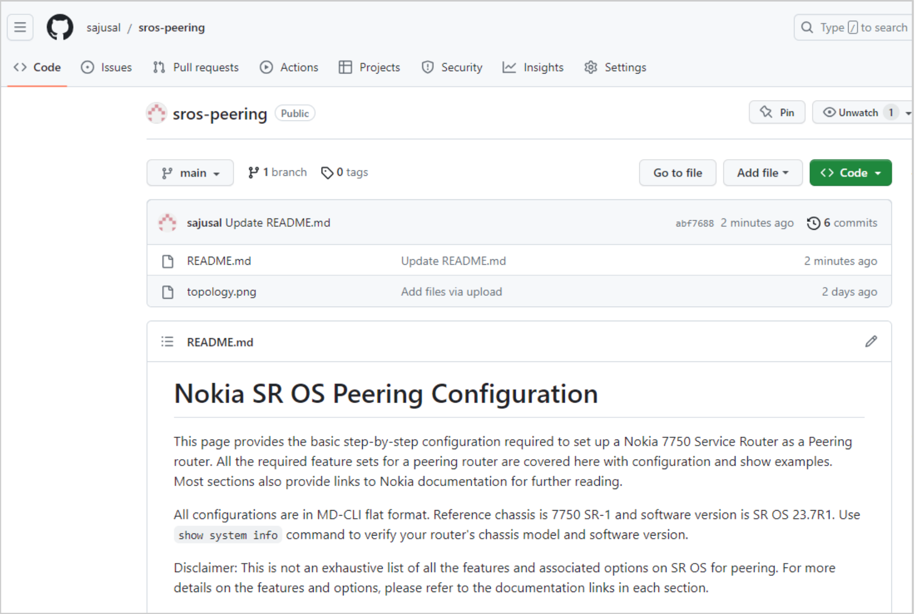
	Linerate	Packet Mix	Ambient Temp	Optics
Vendor A	50%	iMix*	25C	Yes – 12W (400G), subtracted from overall power
Vendor B	15%	768B	23C	No – DAC cables, minimal fan power
Vendor C	50%	iMix*	25C	Yes – optics power subtracted from overall power
Vendor D	50%	iMix*	25C	No – DAC cables, minimal fan power

Real world benchmarks/measurements are critical!

* iMix used is different for all vendors – this is not public info

Peering reference configuration

https://github.com/sajusal/sros-peering


sajusal / sros-peering

Type / to search

Code Issues Pull requests Actions Projects Security Insights Settings

sros-peering Public

Pin Unwatch 1

main 1 branch 0 tags Go to file Add file Code

sajusal Update README.md abf7688 2 minutes ago 6 commits

README.md	Update README.md	2 minutes ago
topology.png	Add files via upload	2 days ago

README.md

Nokia SR OS Peering Configuration

This page provides the basic step-by-step configuration required to set up a Nokia 7750 Service Router as a Peering router. All the required feature sets for a peering router are covered here with configuration and show examples. Most sections also provide links to Nokia documentation for further reading.

All configurations are in MD-CLI flat format. Reference chassis is 7750 SR-1 and software version is SR OS 23.7R1. Use `show system info` command to verify your router's chassis model and software version.

Disclaimer: This is not an exhaustive list of all the features and associated options on SR OS for peering. For more details on the features and options, please refer to the documentation links in each section.

Route Policies

Routing policies control the size and content of the routing tables, the routes that are advertised, and the best route to take to reach a destination.

For more details on route policy configuration and options, visit SR OS Route Policies Documentation.

In these examples, we are creating AS path lists, community and prefix lists.

```
/configure policy-options as-path "PEERING" { expression "64503" }
/configure policy-options as-path-group "BOGON" { entry 10 expression ".* 0 .*" }
/configure policy-options as-path-group "BOGON" { entry 20 expression ".* [64496-64511] .*" }
/configure policy-options as-path-group "BOGON" { entry 30 expression ".* 65535 .*" }

/configure policy-options community "LARGE-PEER" { member "65100:100" }
/configure policy-options community "SMALL-PEERS" { member "65200:200" }
/configure policy-options community "SMALL-PEERS" { member "65400:.*$" }
/configure policy-options community "SMALL-PEERS" { member "65500:.*" }
```

RPKI

SR OS supports RPKI for BGP prefix origin validation.

In this example, we are configuring a RPKI session with an external server and then enabling prefix origin validation under the BGP group. We are also configuring BGP to not use any routes whose origin is invalid.

For more details on RPKI implementation, visit SR OS RPKI Documentation.

```
/configure router "Base" origin-validation rpki-session 172.31.1.2 { admin-state enable }
/configure router "Base" origin-validation rpki-session 172.31.1.2 { local-address 10.10.1.4 }
/configure router "Base" origin-validation rpki-session 172.31.1.2 { port 8282 }
/configure router "Base" bgp group "eBGP-Peering" { origin-validation ipv4 true }
/configure router "Base" bgp group "eBGP-Peering" { origin-validation ipv6 true }
/configure router "Base" bgp best-path-selection { origin-invalid-unusable true }
```

NOKIA

Copyright and confidentiality