Juniper Node Slicing - An experiment in router virtualization for R&E networks

ARE-ON Arkansas Research and Education Optical Network · INTERNET2 · KanREN CONNECTING BEYOND THE CONNECTION · LEARN LONESTAR EDUCATION AND RESEARCH NETWORK · LSU LOUISIANA STATE UNIVERSITY · OneNet Advancing Technology Across Oklahoma

Node Slicing Collaboration

- Retrospective / History
- Community Team
- Introduction to Node Slicing (logical systems ++)
- Use cases for node slicing
- Components required (vs recommended by Juniper)
- The Abstracted Fabric interface
- Node Slicing working group test scenario
- Findings of the Working Group
- Considerations on the economic benefits of node slicing
- Considerations on the operational models of node slicing

What is Network Virtualization?

https://en.wikipedia.org/wiki/Network_virtualization

In computing, **network virtualization** or **network virtualisation** is the process of combining hardware and software network resources and network functionality into a single, software-based administrative entity, a **virtual network**. Network virtualization involves platform virtualization, often combined with resource virtualization.

Network virtualization is categorized as either **external virtualization**, combining many networks or parts of networks into a virtual unit, or **internal virtualization**, providing network-like functionality to software containers on a single network server.

In the networking space, we further look at the terms "Hardware Isolated Router Virtualization" (HVR), and Software Isolated Router Virtualization (SVR).

Juniper Node Slicing - An experiment in router virtualization for R&E networks

What is Network Virtualization?

Category	Hardware-Isolated Virtual Router	Software-Isolated Virtual Router
Control plane resources (CPU, memory)	Dedicated	Shared
Data plane resources (forwarding engine, queues)	Dedicated	Shared
Chassis resources (power supplies, blowers, fabric)	Shared	Shared
Management, configuration	Dedicated	Typically shared, but varies depending on degree of virtualization
Connections between virtualized routing entities	Typically external	Typically internal, but possibly external
Per-chassis scalability (routing adjacencies, prefixes)	Increased with additional logical routers	Unaffected by additional virtual routers

Examples of Vendor Network Virtualization

Juniper

- Logical systems (Software-Isolated Virtual Router)
    - Multiple virtual routers running under the same operating system
    - Separate RPD process for each logical instance
    - CPU / Memory / Line Cards shared between multiple systems
- Node Slicing (Hardware and Software Hybrid Virtualization)
    - Main operating system (BSYS) acts as the hypervisor layer, in control of all PFE resources
    - Guest Network Functions (GNF) run as full running copies (VMs) of the JunOS system alongside the BSYS. Can either run on dedicated external hypervisor (KVM) or on newest routing engines (Max 3 GNF plus BSYS)
    - CPU and memory are fully dedicated to each GNF. Line cards are individually assigned to each GNF = full isolation between GNFs. (integration at backplane).

Examples of Vendor Network Virtualization

Cisco

- Secure Domain Router
  - Available for some time (at least 2008)
  - Requires dedicating Route Processor to each SDR
  - ASR9000 is only a single shelf system with RP/DRP, only supports 1 SDR instance
  - CRS-1 supports 8 SDR instances
- NCS Platform introduced vm based SDR
  - NCS-6008 can support up to 3 SDR instances.
  - VM's run on the route processor, cannot be extended to run on external hypervisor*

Examples of Vendor Network Virtualization

Arista

- Buy more boxes (yes, they really said that to us!)

Feedback from the Internet2 community - Leverage common resources when possible.

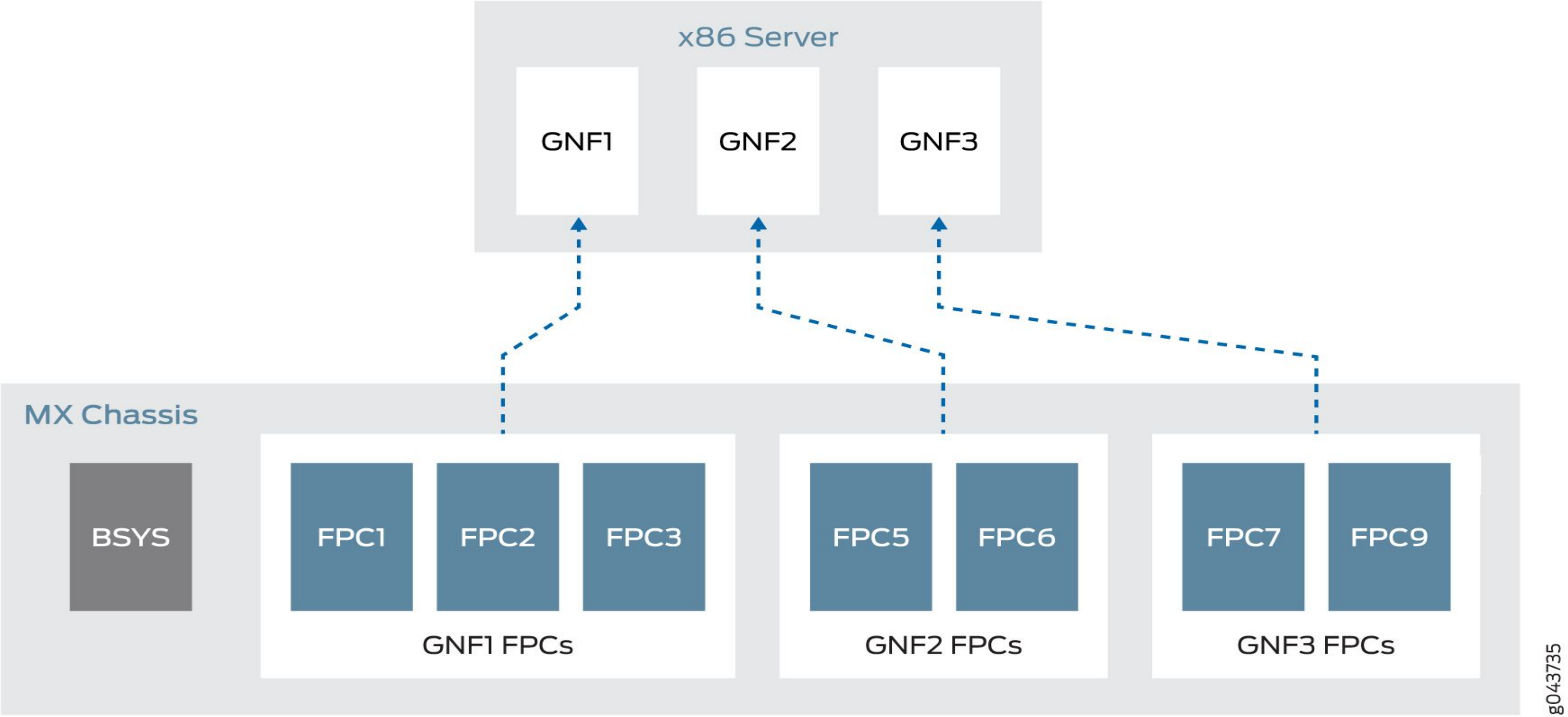
- Internet2 approached by several members who we share the same colo environments
- Rackspace/Power are large operational expenses for organizations
- Can we make better use of our footprints to save $$$?
- Committee formed by several leaders in the R&E Community - Akbar Kara (LEARN), James Deaton (GPN)
- Maybe Internet2 can leverage routers owned by other institutions, or vice versa?
- JunOS is very common in the R&E community, so Node Slicing seemed like the obvious choice.
- Several engineers from other organizations contributing to the testing - Brad Fleming (GPN), Andrew Laubach (OneNet), Bobby Clark (OneNet), Robert Nordmark (OneNet), Byron Hicks (LEARN), Matt Mullins (GlobalNOC/Internet2)

Juniper Node Slicing (logical systems++)

- Multiple partitions in a single physical MX Series router - Guest Network Functions (GNFs).
- Each GNF behaves as an independent router, with its own dedicated control plane, data plane, and management plane. This enables you to run multiple services on a single converged MX Series router, while still maintaining operational isolation between them.
- Partitions that do not share the control plane or the forwarding plane, but only share the same chassis, space, and power.
- You can also send traffic between GNFs directly through the switch fabric - Also known as the Abstracted Fabric or "AF" Interface.
- GNF's typically run as VM's on an external physical server, orchestrated by a piece of software called the Juniper Device Manager (JDM)

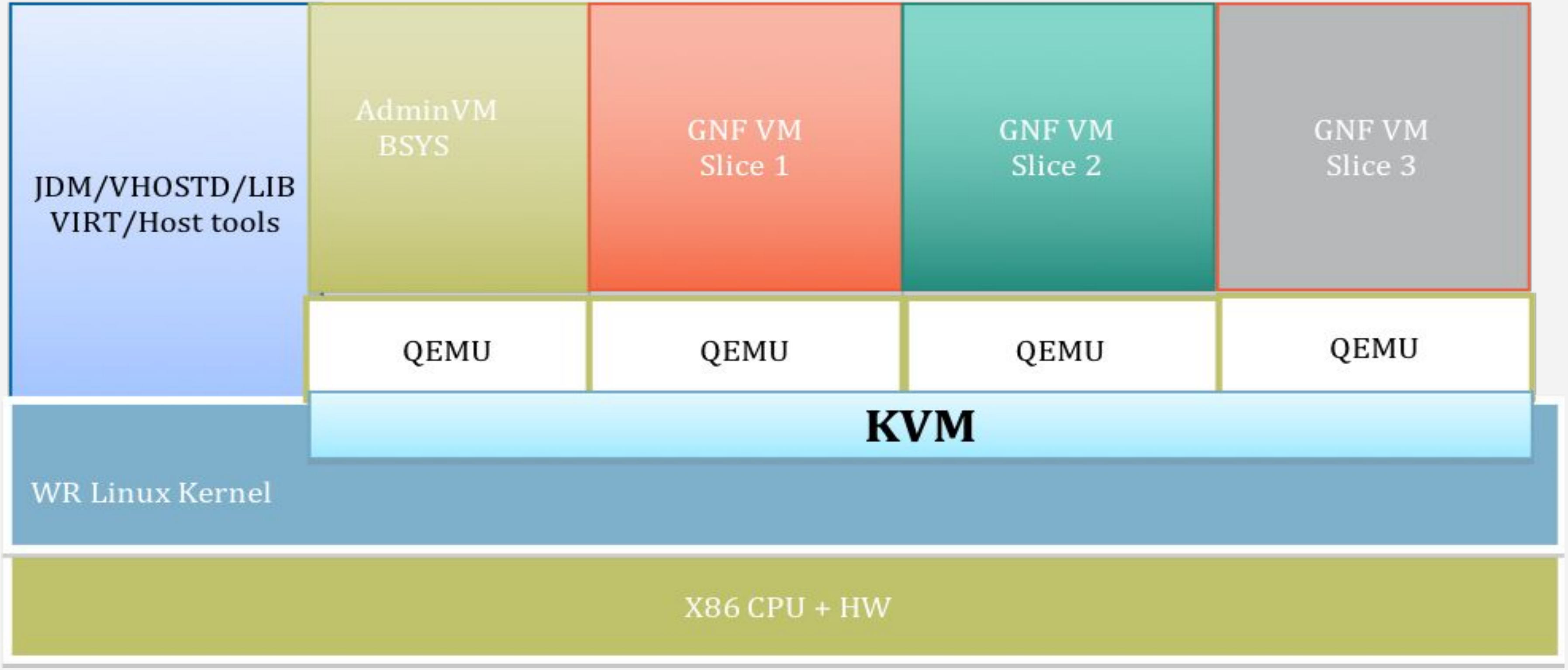
Juniper Node Slicing - An experiment in router virtualization for R&E networks

Juniper Node Slicing (logical systems++)



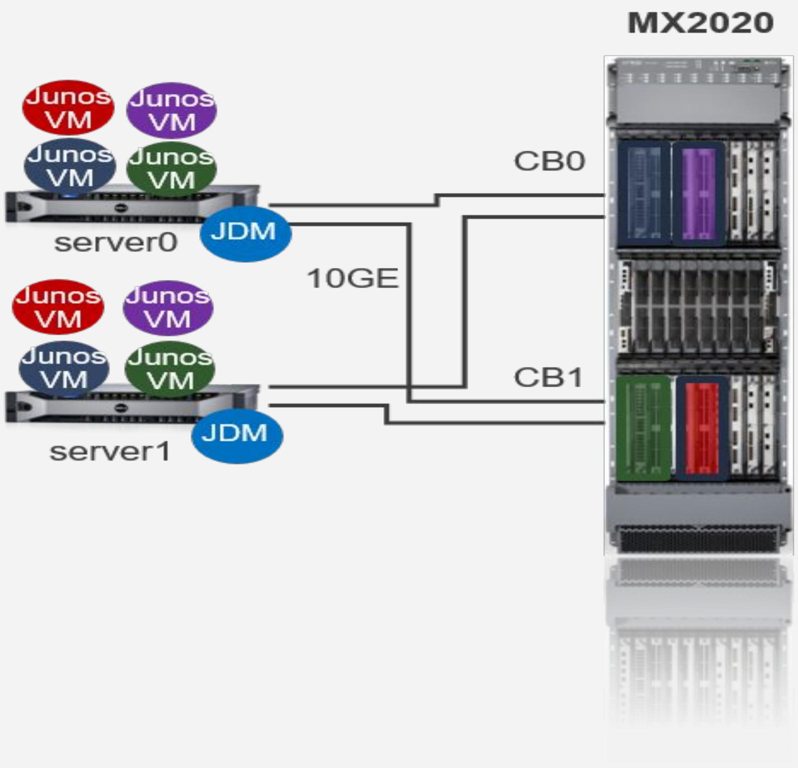
Juniper Node Slicing - An experiment in router virtualization for R&E networks

Juniper Node Slicing (logical systems++)



Juniper Node Slicing - An experiment in router virtualization for R&E networks

Typical Connectivity - Full Redundancy of Control Plane / SCB



Juniper Node Slicing - An experiment in router virtualization for R&E networks

Configuration - Juniper Device Manager (JDM)

```
root@jdm-server0> show configuration virtual-network-functions
kanren {
    id 5;
    chassis-type mx480;
    resource-template 2core-16g;
}
Internet2 {
    id 4;
    chassis-type mx480;
    resource-template 2core-16g;
}
OneNet {
    id 1;
    chassis-type mx480;
    resource-template 2core-16g;
}
```

```
request virtual-network-functions add-image /var/jdm-usr/gnf-images/junos-install-ns-mx-x86-64-17.4R1.16.tgz
<name-of-gnf> all-servers
```

Configuration - Router BSYS

```
rtrslc@NODE-SLICING-RE0> show configuration chassis network-slices
guest-network-functions {
    gnf 1 {
        description OneNet;
        fpcs [ 2 3 ];
    }
    gnf 4 {
        description Internet2;
        fpcs [ 0 4 ];
        af1 {
            peer-gnf id 2 af1;
        }
        af2 {
            peer-gnf id 3 af2;
        }
        af3 {
            peer-gnf id 5 af1;
        }
    }
..
```

GNF INTERCONNECTION WITH AF INTERFACES

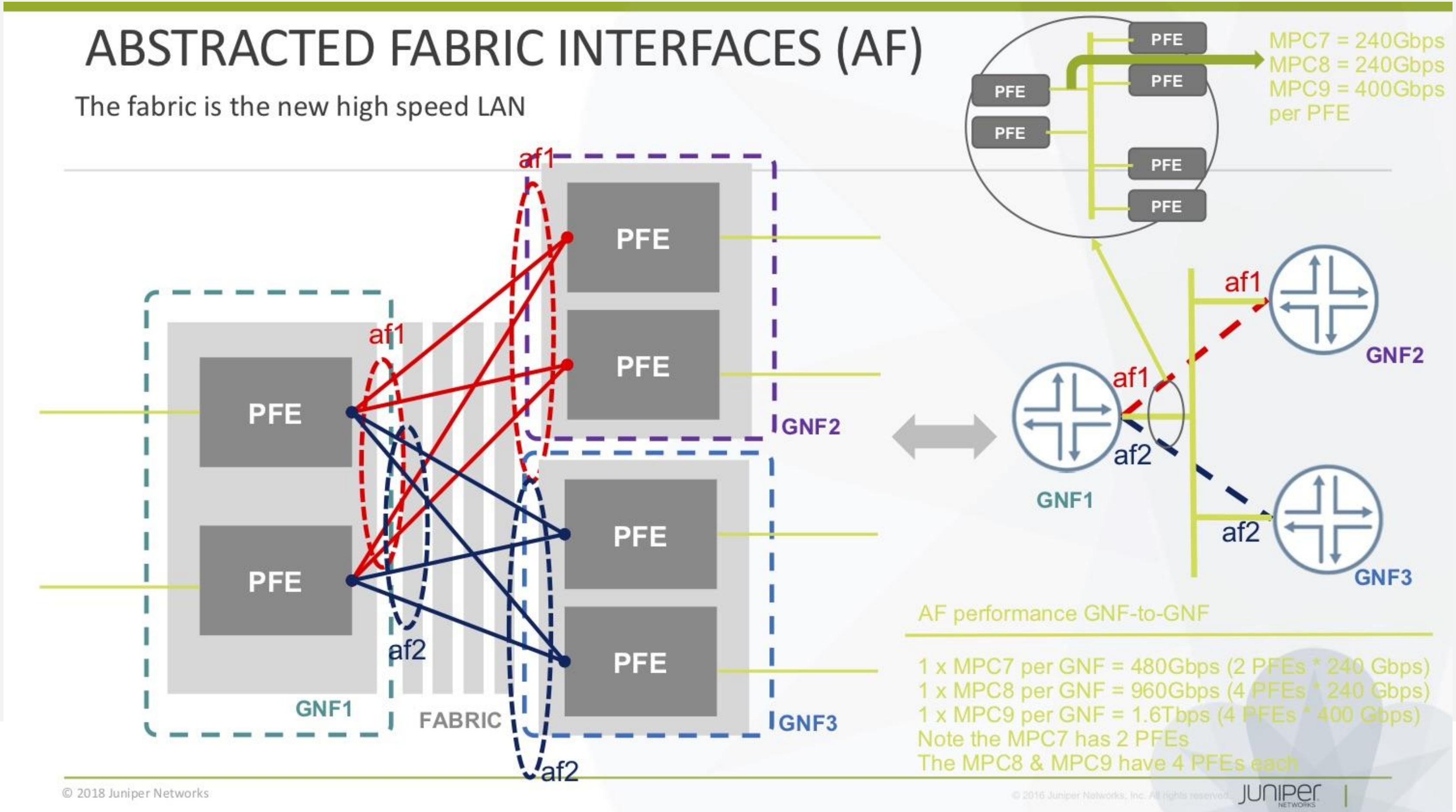
Simplicity at its best

AF configuration at BSYS

```
network-slices {
        guest-network-functions {
            gnf 1 {
                fpcs 0;
                af2 {
                    peer-gnf id 2 af1;
                }
            }
            gnf 2 {
                fpcs 3;
                af1 {
                    peer-gnf id 1 af2;
                }
            }
        }
}
```

"Plug a cable between GNF1 and GNF2"

Juniper Node Slicing - An experiment in router virtualization for R&E networks

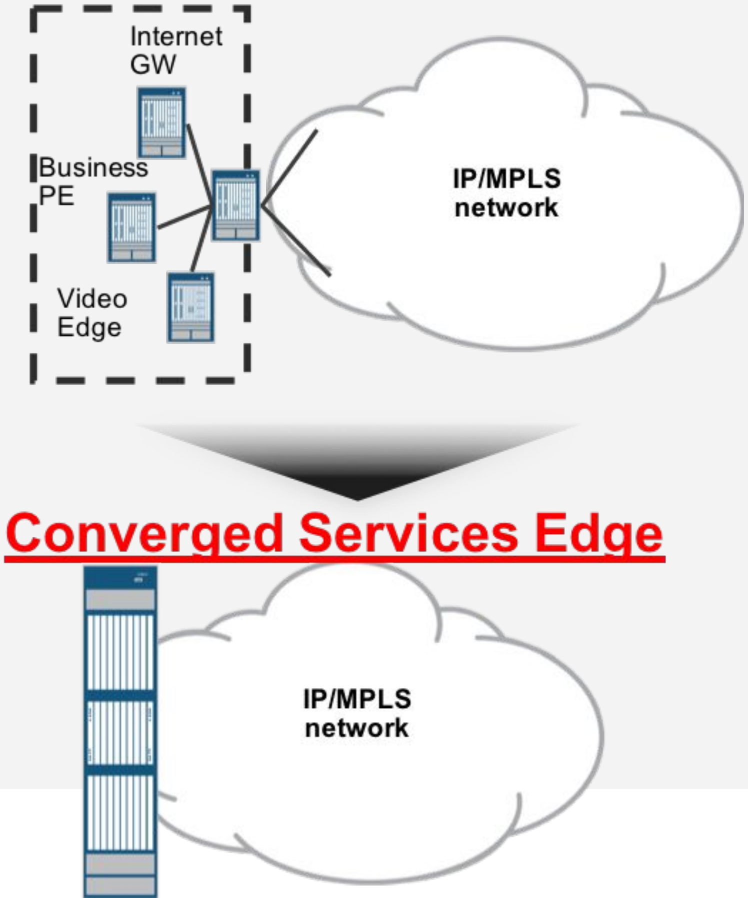


Use Cases for Node Slicing

- Convergence - Reducing CapEx and OpEx by combining multiple network functions (BNG, PE, video edge, voice edge, etc.) into a single device.
- Risk management - Partitioning physical systems, such as BNG, into multiple independent logical instances so that a single instance failure would not result in a massive service outage.
- Device consolidation - Uniting network functions within a single fully redundant physical device on a distributed infrastructure. Business VPN PEs are a prime example of how consolidation significantly reduces OpEx while sustaining existing SLAs.
- Quicker Service Delivery Time - Add new nodes with dedicated services while reducing the risk of configuration error on preexisting service nodes
- Network as software - Treating the network like virtualization and dockers have changed the systems world.

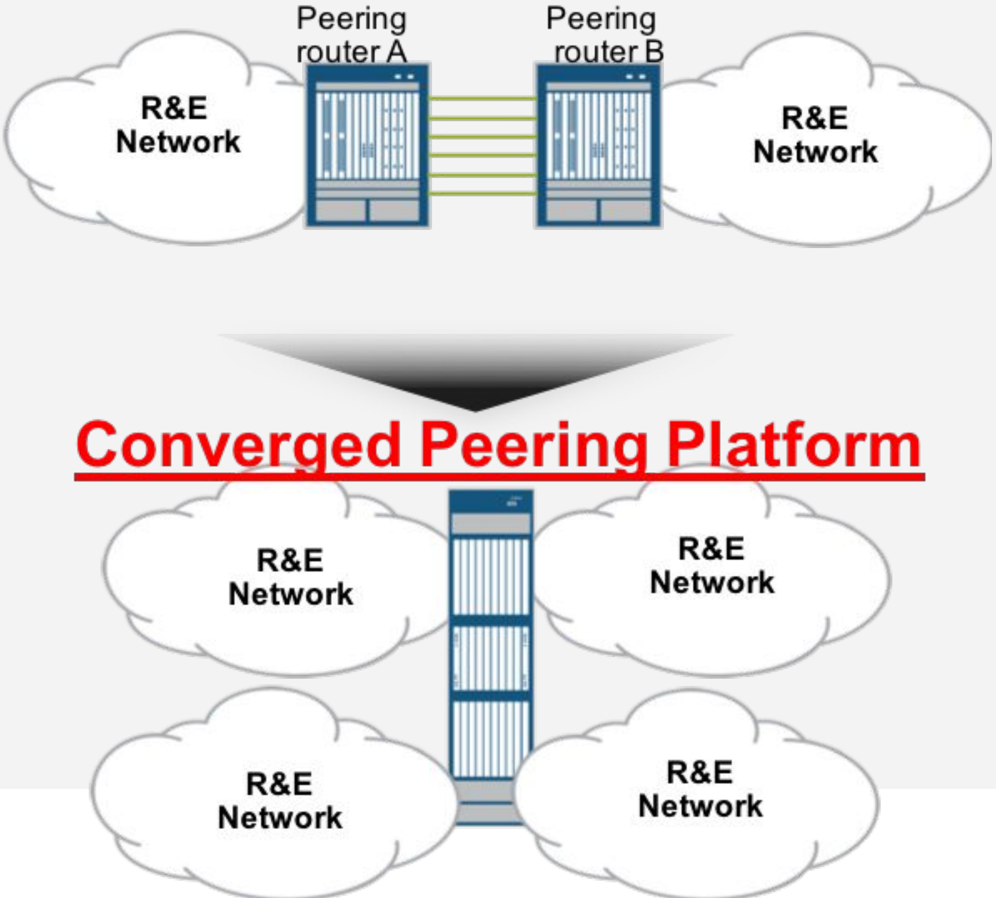
Juniper Node Slicing - An experiment in router virtualization for R&E networks

Use Cases for Node Slicing



Juniper Node Slicing - An experiment in router virtualization for R&E networks

Use Cases for Node Slicing - contd



Juniper Node Slicing - An experiment in router virtualization for R&E networks

Components Required - Per Juniper Documentation

Chassis: MX480, MX960, MX2010, MX2020, and more to come.
Switch Control Board: SCBE2
X86 Server: Num CPUS: Dependent on number of GNF's deployed. A GNF can use one of 4 basic resource templates: 2 Core/16G RAM, 4 Core, 32G RAM, 6 Core, 48G RAM, 8 Core, 64G RAM.
The CPUs must be dedicated to a GNF.
Each GNF requires a minimum of 64 GB storage. (SSD)

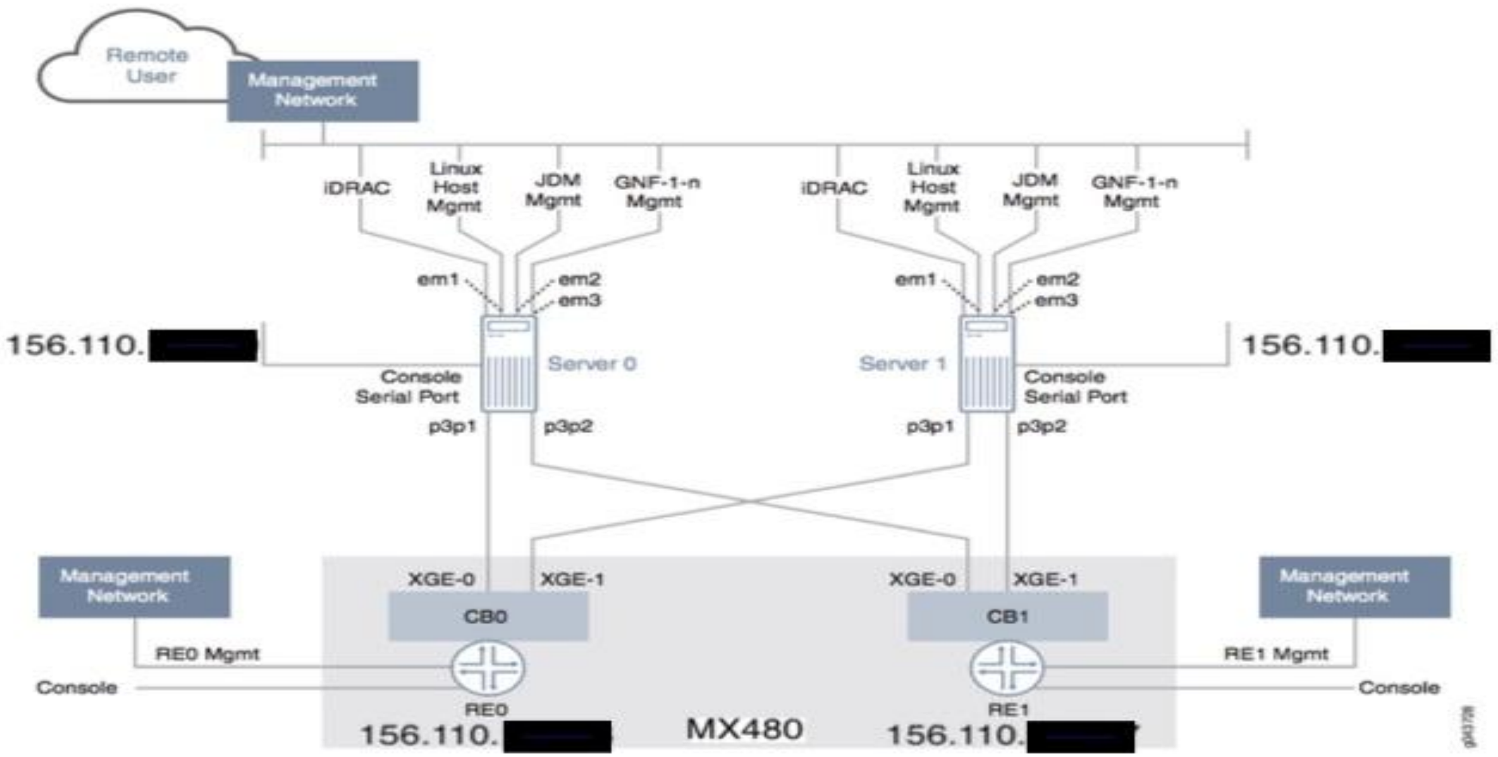
In addition to the GNF requirements, the server must support the following for the shared resources (JDM and Linux host)

- CPU - Four cores to be allocated for JDM and Linux host processing. (Haswell-EP minimum req'd, https://www.juniper.net/documentation/en_US/junos/information-products/pathway-pages/junos-node-slicing/junos-node-slicing.pdf p16)
- Memory - Minimum of 32 GB DRAM for JDM and Linux host processing
- Storage - Minimum of 64 GB storage for JDM and Linux host.
- Network/Comms -
  - Two 10-Gbps Ethernet interfaces for control plane connection between the server and the router.
  - Minimum—1 PCIe NIC card with Intel X710 dual port 10-Gbps Direct Attach, SFP+, Converged Network Adapter, PCIe 3.0, x8
  - Recommended—2 NIC cards of the above type. Use one port from each card to provide redundancy at the card level.
  - One Ethernet interface (1/10 Gbps) for Linux host management network.
  - One Ethernet interface (1/10 Gbps) for JDM management network.
  - One Ethernet interface (1/10 Gbps) for GNF management network. (This port is shared by all the GNFs on that server).
  - Serial port or an equivalent interface (iDRAC, IPMI) for server console access.

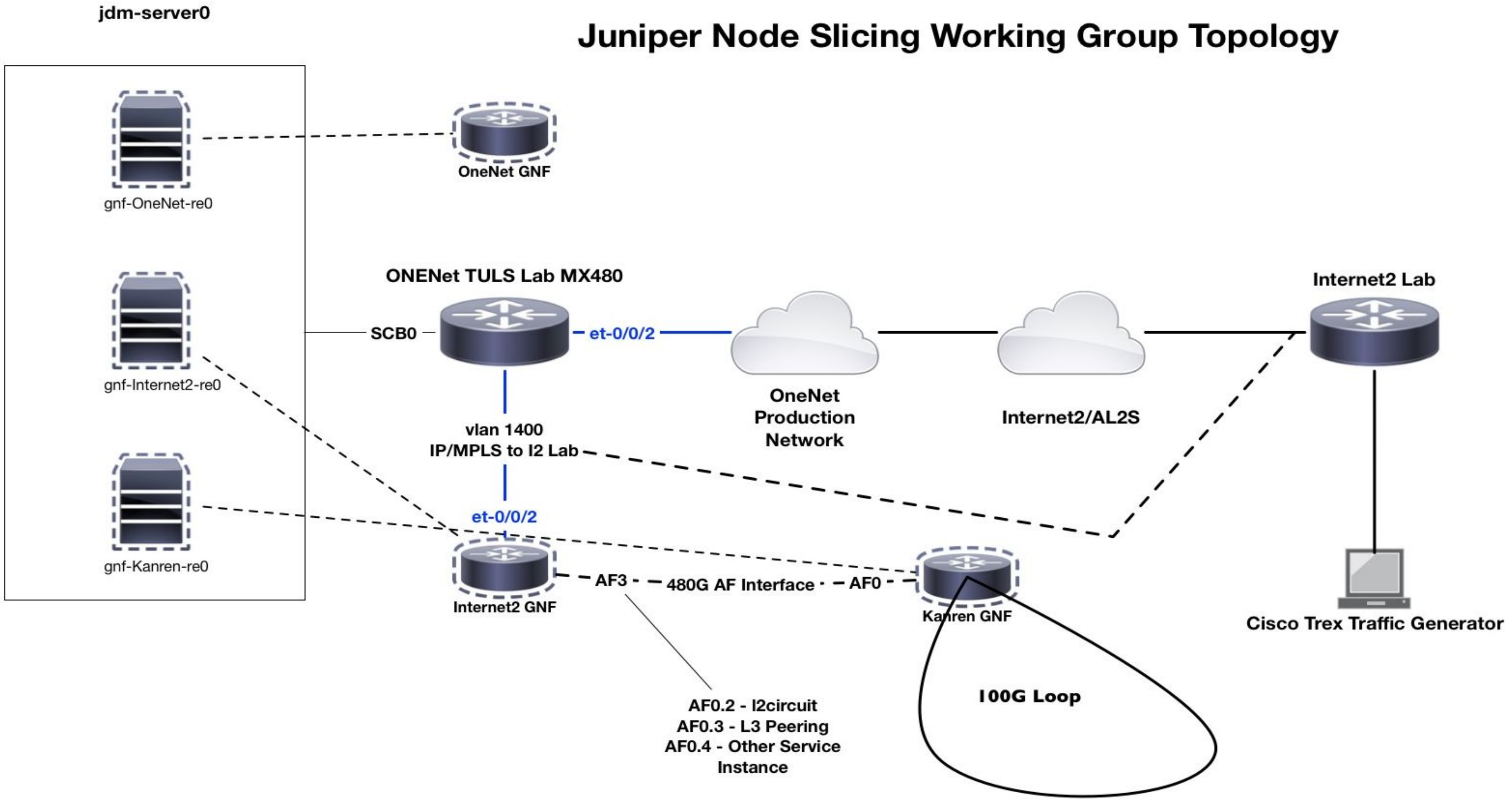
The host OS must be Redhat Enterprise Linux or Ubuntu Linux running KVM.

Juniper Node Slicing - An experiment in router virtualization for R&E networks

OneNet Node Slicing Lab setup



Juniper Node Slicing Working Group Topology



Juniper Node Slicing - An experiment in router virtualization for R&E networks

What we tested

- Provisioning - Straightforward
- Routing Engine Failover - Works
- Route Convergence - Good - Most likely depends on CPU of JDM Server
- Feature Parity of “AF” Interface - Some gotchas

AF Interface Tests

Flexible-vlan-tagging (q-in-q) - brad working

Flexible-ethernet-services - Works but severely limited - I2

IP/MPLS All Types/Signaling - OSPF/ISIS - Works - I2

Inline Flow monitoring / IPFIX / SFLOW - I2 - Not Working (inline jflow) did not try SFLOW

VLAN push/pop/swap operations support (AF ints) - brad not working

QoS/DSCP - Brad - kinda.. See “class of service feature summary”

LLDP - Not working - I2

BFD - brad working

URPF - Works

VPLS - I2/Brad - Not supported at the interface level nor the vlan level

L2Circuit/L2vpn - Working

L3 Service via L3VPN - Working

Juniper Node Slicing - An experiment in router virtualization for R&E networks

Per Juniper:

AF interfaces support the following features:

- Load balancing based on the remote GNF line cards present
- Class of service (CoS) support:
  - Inet-precedence classifier and rewrite
  - DSCP classifier and rewrite
  - MPLS EXP classifier and rewrite
  - DSCP v6 classifier and rewrite for IP v6 traffic
- Support for OSPF, IS-IS, BGP, OSPFv3 protocols, and L3VPN
- Multicast forwarding
- Graceful Routing Engine switchover (GRES)
- MPLS applications where the AF interface acts as a core interface (L3VPN, VPLS, L2VPN, L2CKT, EVPN, and IP over MPLS)

The following protocol families are supported:

- IPv4 Forwarding
- IPv6 Forwarding
- MPLS
- ISO
- CCC

NOTE: The non-AF interfaces support all the protocols that work on Junos OS.

AF Interface rule of thumb:

If the feature depends on any kind of bridging (VPLS, EVPN, Bridge Domain), it won't function.

If you plan on using node slicing in the core of your network, this probably won't be an issue as the AF Interface will most likely be backbone facing.

However, if you plan on terminating services on the AF Interface, your options will be limited to vanilla Layer3 plus some additional edge MPLS functionality (unsupported)

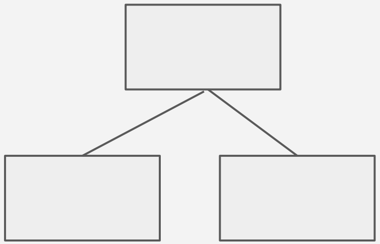
Economic Analysis - Node Virtualization vs Multiple Boxes

- Capex of 1 "Large" Chassis and Routing Engines compared with multiple boxes (this depends highly on how much FIB Scale and Buffering needed)
- Space and Power consumption of multiple boxes. For example, 4 MX10003 Routers vs single MX960/10008
- Opex saved by leveraging node slicing (operational/engineering)
- Cost of interconnecting multiple boxes with external interfaces vs using AF Interfaces. Keep end mind that for an AF interface to function, it needs to be bound to MPC7 or higher "NG" card.
- Cost of the GNF feature licence.
- Cost of the AF Interface feature licence.
- Capex/Opex of JDM servers
- Space/Power taken up by JDM servers (this can be offset by running the GNF on the RE-NG, but unsure if this is fully supported)

Juniper Node Slicing - An experiment in router virtualization for R&E networks

Part Number	Description	List Price
JNS-AF-100G	Abstracted Fabric (AF) license for 100Gbps. Perpetual license	$30,000.00
JNS-AF-10G	Abstracted Fabric (AF) license for 10Gbps. Perpetual license	$3,000.00
JNS-AF-400G	Abstracted Fabric (AF) license for 400Gbps. Perpetual license	$100,000.00
JNS-AF-40G	Abstracted Fabric (AF) license for 40Gbps. Perpetual license	$12,000.00
JNS-GNF-EXT	Perpetual Software License for a Guest Network Function (GNF) where the Junos VM runs on a External x86 Server	$15,000.00
JNS-GNF-EXT-1Y	1 Year Subscription License for one Guest Network Function (GNF) on MX where the Junos Control Plane VM runs on a External x86 Server	$5,930.00
SVC-COR-JNS-A-100G	Juniper Care Core Support for JNS-AF-100G	$900.00
SVC-COR-JNS-A-10G	Juniper Care Core Support for JNS-AF-10G	$90.00
SVC-COR-JNS-A-400G	Juniper Care Core Support for JNS-AF-400G	$3,000.00
SVC-COR-JNS-A-40G	Juniper Care Core Support for JNS-AF-40G	$900.00
SVC-COR-JNS-GNF-EX	Juniper Care Core Support for JNS-GNF-EXT	$450.00

Juniper Node Slicing - An experiment in router virtualization for R&E networks



Operational Concerns

Juniper expects the GNF machines to be running a codetrain no more than 2 major releases ahead or behind. For example If the BSYS is running 19.2, then a GNF can be in a range within 18.4 and 19.4. This model means that all GNFs need to run very similar codetrains, perhaps this might not work for all “tenants”.

Many people in the R&E community use Juniper, but for those that don’t?

Can tenants open JTAC cases on issues with their GNF if the JDM/chassis is operated by another party?

Providers need to establish a pricing model to share the hardware out to their tenants. Perhaps the tenants each pay for a line card, but what about the underlying chassis and components?

Standard trust model exists where tenants trust their providers to ensure the control plane (JDM) is stable and secure. Many organizations have strict security requirements on infrastructure security that need to be met, can providers meet these standards?

Cloud providers have solved these issues but it adds some additional complexity.

Internet2 Specific Issues

Limitations on AF interfaces make it challenging for us as a service provider to use these as service terminating interfaces.

- Lack of vlan manipulation presents a challenge since many members translate vlans on either sides of their circuits, especially in cloud connect scenarios
- Lack of VPLS and EVPN/Bridging presents an issue. Workarounds exist (Stretch Attachment circuit to a bridging capable node, or use family CCC and connect into a Psedowire Subscriber (PS) interface. However, these are hacks and scaling isn't clear.
- Lack of per-unit scheduling isn't currently an issue, but if QoS sees more adoption in R&E, it could be.
- Biggest issue is lack of support for inline flow. Analytics/Telemetry are critical for us to optimize peering.

According to Juniper, hopefully all these issues can be addressed in 19.4 (2H2019) with enough customer feedback. Please contact your SE/Account Rep if these issues matter to you!

The Takeaway

Node slicing is a viable technology for production scenarios

The lack of features on AF interfaces (inter-GNF) and the lack of per-pfe apportioning (line card sharing) prevents certain adopters (provider->tenant) from maximizing the economic value. This may be a somewhat unique case for R&E environments - where the provider - tenant relationship is fairly close.

Economic benefits must be evaluated on a case by case basis; they may be further challenged by commercial development of smaller chassis (1U in some cases) with n x 100G interfaces

Provides vlear value for organizations that have separate administrative areas (ie Security/Network) that are currently running multiple boxes to reduce risk

Special Thanks to the following contributors:

- Brad Flemming (KanRen) - Huge amount of testing of features with I2 as well as within his own lab.
- Andrew Laubach, Robert Nordmark (OneNet) - For setting up the Node Slicing Lab, and plumbing various parts as needed
- Matt Mullins (Internet2 GlobalNOC) - Helping with feature validation for Internet2
- Bobby Clark, Byron Hicks, David Merrifield, Gary Mumphrey - Community Contributors
- Akbar Kara (LEARN), Robert Nordmark (OneNet), Chris Wilkinson (Internet2) - Leadership/Organizational Support
- Von Royal (OneNet), Pankaj Shah (LEARN), Rob Vietzke (Internet2) - Community Steering Team