HEPiX Network Virtualisation WG Update

Marian Babik, Shawn McKee

GDB, FNAL, Sept 2019

Introduction

- High Energy Physics (HEP) has significantly benefited from strong relationship with Research and Education (R&E) network providers
  - Thanks to LHCOPN/LHCONE community and NREN contributions, experiments enjoy almost “infinite” capacity at relatively low (or no-direct) cost
  - NRENs have been able to continually expand their capacities to overprovision the networks relative to the experiments needs and use
- Other data intensive sciences are coming online soon (SKA, LSST, etc.)
  - Besides Astronomy there are MANY science domains anticipating data scales beyond LHC
- Network provisioning will need to evolve
  - Focusing not only on network capacity, but also on other network capabilities
- It’s important that we explore new technologies and evaluate how they could be useful to our future computing models
  - While it’s still unclear which technologies will become mainstream, it’s already clear that software (software-defined) will play major role in networks in the mid-term

Network Functions Virtualisation WG

Mandate: Identify use cases, survey existing approaches and evaluate whether and how Software Defined Networking (SDN) and Network Functions Virtualisation (NFV) should be deployed in HEP.

Team: 60 members including R&Es (GEANT, ESNet, Internet2, AARNet, Canarie, SURFNet, GARR, JISC, RENATER, NORDUnet) and sites (ASGC, PIC, BNL, CNAF, CERN, KIAE, FIU, AGLT2, Caltech, DESY, IHEP, Nikhef)

Monthly **meetings** started last year (https://indico.cern.ch/category/10031/)
Mailing list: https://listserv.in2p3.fr/cgi-bin/wa?SUBED1=hepix-nfv-wg

Objectives/sub-tasks

- Work organised in **two phases**, phase I (exploratory):
  - Detail use cases relevant for HEP
    - Explore SDN/NFV approaches for compute, e.g. OpenStack/Kubernetes (mainly **intra-site** activities) - **Cloud Native DC Networking**
    - Explore SDN/NFV approaches for distributed storage/end-to-end transfers, e.g. data lakes (**inter-site** activity in collaboration with RENs/NRENs) - **Programmable Wide Area Networks**
  - Tutorials/introductory material to help sites establish their testbeds
  - Evaluate existing approaches, analyze readiness/gaps
  - Promote sharing of experiences between the sites/RENs/NRENs
  - Document deployment experiences, issues/gaps, production readiness
- WG Report - in progress - draft will be available by HEPiX fall
  - If there will be sufficient momentum/effort for phase II then:
    - Propose timetable and analyse resource needed to run cross-site experiments/testbeds
    - Implementation and configuration advice, organise scalability/performance testing

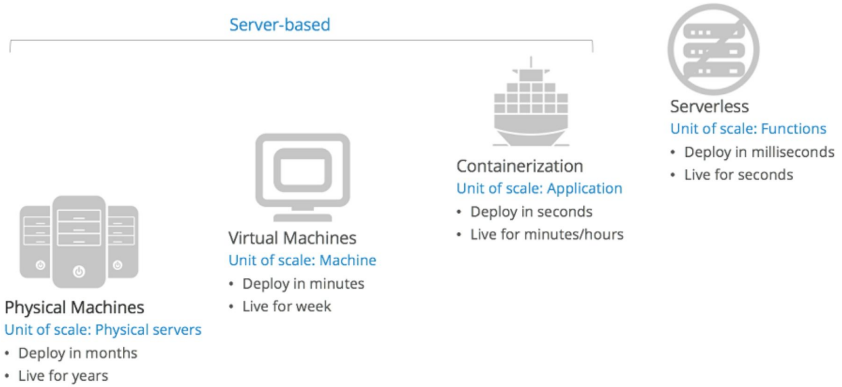
Cloud Native DC Networking

Paradigm Shift in Computing

Moving from static physical machines to very dynamic models with VMs, containers, clusters of containers and federated clusters/serverless

This has major impact on networking requirements in DC

- Nodes can appear/disappear in msecs
- East-west traffic increases
- Nodes can migrate (even across DCs)
- Multiple orchestration methods (stacks) need to co-exist in the DC network
- Networking across stacks and within needs to perform



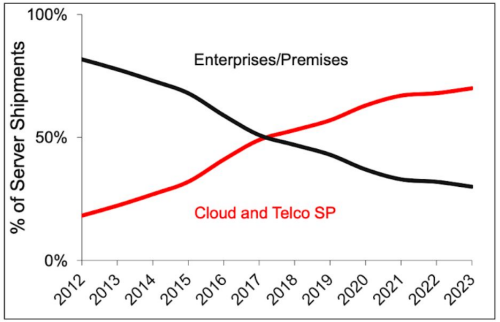
This transition has already started, we already have experiments running payloads in containers and services bundled in K8s pods, physics analysis in K8s has been demoed recently

Cloud Network Opportunity

Paradigm shift most notable in compute, but networking evolution is also pushed by virtualized storages and GPUs (wrt. expected throughput and latency)

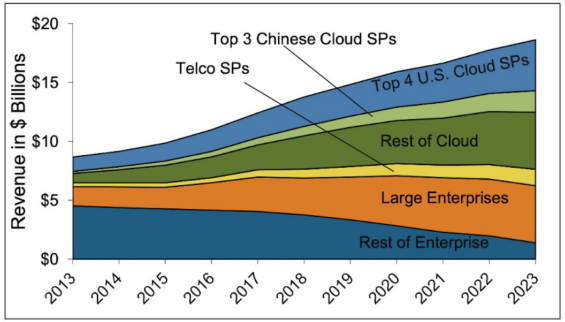
Network vendors have already started to take note - this will impact future HW

Server Shipments



Source: Dell'Oro Group Server Research

Data Center Ethernet Switch Revenue



Source: Dell'Oro Ethernet Switch Data Center 5 Year Forecast Report January 2019

Enterprise workloads are migrating to clouds and hybrid clouds

Emergence of cloud native apps and containers necessitates new architecture

Use Cases

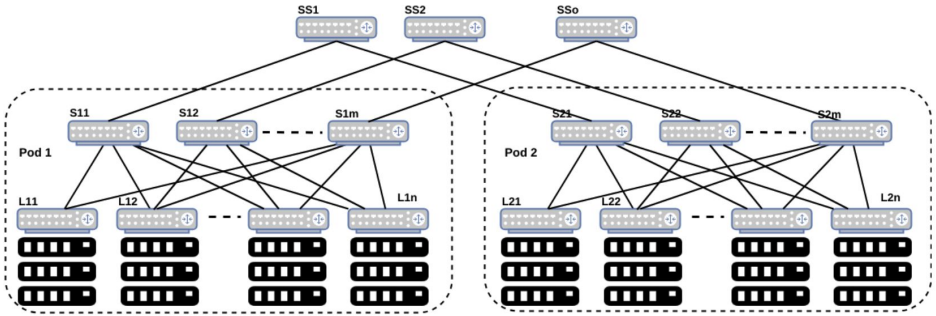
Data centre networking offering standard cloud compute services

- Native support for multi-stack
  - Connecting and integrating multiple orchestration stacks like k8s, OpenStack, etc.
  - Networking and security across legacy, virtualized and containerized applications
- Network support across-stack
  - Networking and security across legacy, virtualized and containerized applications
- Native support for multi-cloud
  - Extending DC networks to Commercial Clouds and creating federated services spanning DCs
- Multi-tenancy/isolation
  - Support for application/experiment level networking (e.g., MultiONE presentation earlier)
- Network automation
- Security and observability
  - Multistack and across-stack policy control, visibility and analytics

Evolution in DC networking

Clos-topology now de-facto standard

- Homogenous & simple equipment
- Easy to scale and add capacity
- **Routing is the core interconnect technology**
  - Bridging/switching only at the leafs (within a single rack)
  - Connecting across racks relies on **network virtualisation**



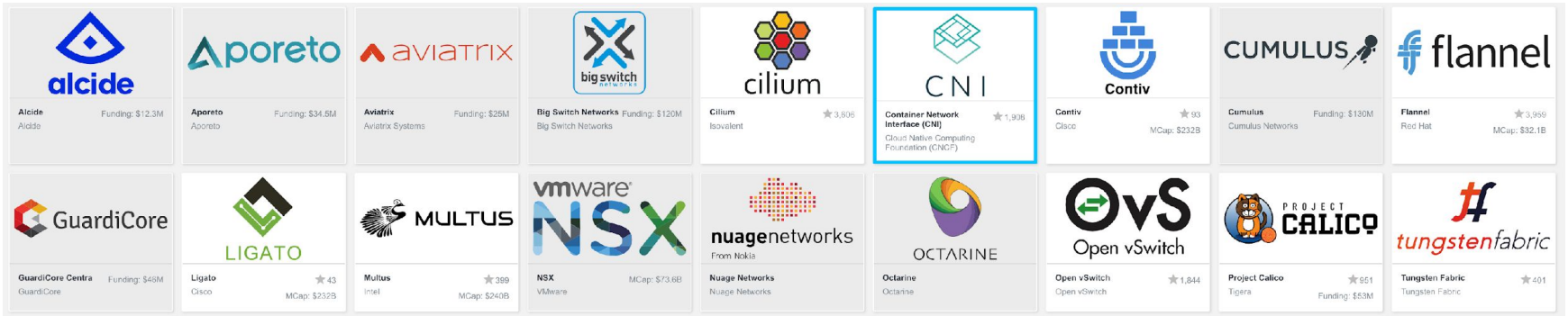
Sample Tier-2 Clos topology, compute is attached to leafs/ToRs, each leaf is connected to multiple spines (altogether forming a pod) and each spine within a pod connects to super-spine.

- Control plane pushed all the way to the leafs (or even directly in servers running software switches)
- Fine-grained failure domain
- Simple homogenous equipment pushes evolution towards open source
  - Both in software as well as hardware equipment
- Clos is a topic of its own, many different topologies/possibilities exist

Network Virtualisation

Carves a single physical network into multiple, isolated virtual networks

Range of approaches, both open-source (white) and commercial (grey) exist, tracked by the [Cloud Native Computing Foundation]



In open source there are currently three different lines of thought:

- OpenFlow - pursued by OVN/OVS/ODL and others
- BGP/MPLS - pursued by Tungsten/Contrail (Linux Foundation project now)
- FRR - pursued by Cumulus

Network Virtualisation - OpenFlow

- OpenFlow started with an influential paper and became a movement in networking R&D
  - The core idea is to use flow tables (available in most packet switching silicon) and use OpenFlow protocol to remotely program the tables (from a centralised controller)
- OpenVSwitch (OVS) is an open source implementation of pure OpenFlow software switch
  - Native controller to program it is Open Virtual Network (OVN) (but others can be used as well)
  - Data plane can use VXLAN, GRE, Geneve; control plane is OpenFlow or native OVSDB
  - Controller supports integration with OpenStack and K8s
- OpenFlow protocol has been updated several times to address its shortcomings and overall didn't live up to its expectations
  - However there are existing production deployments (Google)
  - OpenFlow as such has proven to be very useful in other areas (WAN use cases)
  - Flow tables are still core part of some key network functions (ACLs, NAT, etc.)

Network Virtualisation - BGP/MPLS

An alternative to OpenFlow is to look if it's possible to adapt to DC other protocols that we know well

- As we have run the Internet with them for years, e.g. BGP/MPLS

Tungsten/Contrail is using a combination of MP-BGP and BGP/MPLS (rfc4684):

- It has its own software switch running on servers (vrouter), which implements VRFs and uses combination of MPLS/VXLAN to connect them*
- Supports MPLSoUDP, MPLSoGRE, VXLAN; BGP/EVPN in control plane
- Native integration with physical equipment (though likely best with Juniper**)

Supports both multi-stack and across-stack (OpenStack, VMware, K8s)
Using BGP/MPLS-VPN internally means it's easy to extend network to other DCs
Also enables easy extension to Clouds (by bringing there your own network)

*For comparison btw OVS and vrouter see OVS talk by Y. Yang

Network Virtualisation - FRR

Unlike previous solutions that use software switches, another alternative is to implement core switching/routing functions directly in Linux kernel and create an open source network OS

This approach is followed by Cumulus and others
- sometimes with their own equipment (using merchant packet switching silicon)

Free Range Routing (**FRR**) - IP routing suite for Linux that supports range of routing protocols BGP, OSPF, IS-IS (Linux Foundation project, formerly Quagga).

As this is a platform, different approaches are possible. One approach is to run eBGP as the only control plane in DC and use EVPN/VXLAN to integrate with compute.

Linux Foundation Networking

Additional projects that improve performance, provide alternative controllers, offer programmable off-loading capabilities, etc. are hosted by Linux Foundation

Intel's Data Plane Development Kit (DPDK) - accelerates packet processing workloads running on a wide variety of CPU architectures

P4 - programming language for packet processing - suitable for describing everything from high- performance forwarding ASICs to software switches.

SmartNICs

- Now offered from multiple vendors - goal is to maximise capacity while providing full programmability for virtual switching and routing, tunnelling (VXLAN, MPLS), ACLs and security groups, etc.
- Three approaches are being followed:
  - FPGA based - good performance, but difficult to program, workload specific optimisation
  - ASIC based - best price/performance, easy to program but extensibility limited to pre-defined capabilities
  - SOC based - good price/performance, easily programmable, highest flexibility
- Datapath programmability (tutorial)
  - Application level - OpenVSwitch, Tungsten vRouter, etc.
  - Packet movement infrastructure (part of data path) - BPF (Berkeley Packet Filter)/eBPF
  - Full description of data path - P4 language
- FPGA-based SmartNICs broadly deployed in Microsoft Azure
- Tungest Fabric 5.1 release plans to support smartNICs
- Good overview provided in ACM SIGARCH article

Programmable Networks

Use Cases

Focusing mainly on WAN/**SD-WAN** like capabilities

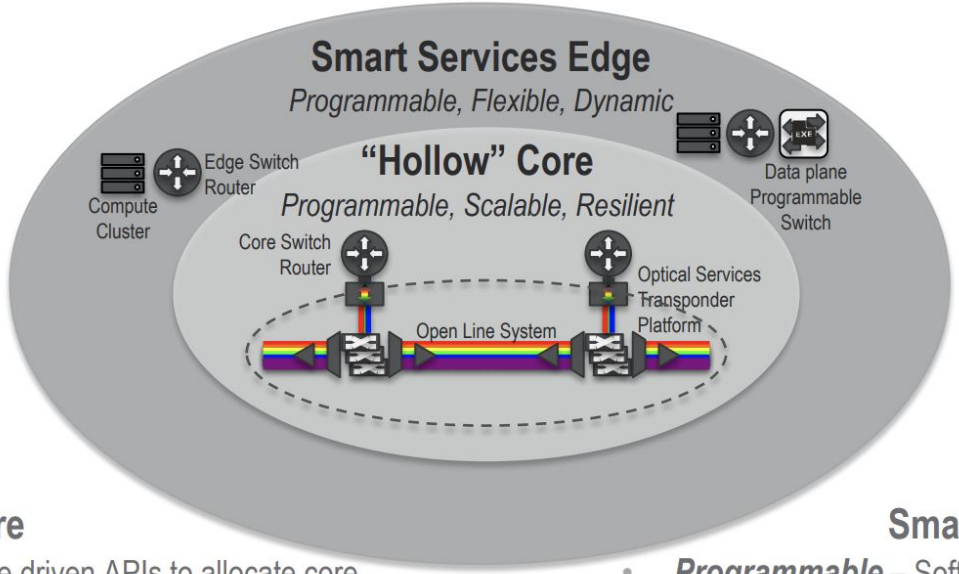
- Traffic engineering
  - Additional capacity exists and can be provisioned by steering traffic via alternate paths
- Network provisioning
  - With DC networking moving towards WAN protocols, there is an opportunity to leverage this to find alternative ways how to organise/manage current L3VPNs/LHCONE
- Provide QoS transfers
  - We have been running two dedicated networks (LHCOPN and LHCONE) which mainly differ in QoS provided. Other experiments will likely come up with similar requirements.
- Improve network to storage performance
  - Currently there is often a mismatch between target storage and network performance
- Capacity sharing
  - Monitoring and managing network as a resource in a similar way we do compute and storage today is becoming likely in the future
- Effective use of HPCs and Clouds

R&E Plans

- R&E network providers have long been working closely with HEP
  - HEP has been representative of the future data intensive science domains
  - Often serving as testbed environment for early prototypes
- Surveying their plans for higher-level services and providing our feedback is critical for future evolution of HEP networking
- Different approaches are being followed - ranging from full SDN capable services (AMLight) up to a range of various low to higher-level edge services (ESNet6)
- Some important questions to address:
  - How do we tackle such range of network capabilities across R&Es ?
  - What services will be offered, how do we accommodate different functionality across R&Es ?
  - What interdependencies exist between Cloud Native Networking and R&E plans ? Can we effectively run DCIs over future networks ?
- **Our ability to use the programmable edge services will directly impact our ability to effectively use future networks.**

ESnet6 ("Hollow-Core") Architecture Overview

Orchestration and Automation



Monitoring and Measurement

"Hollow" Core

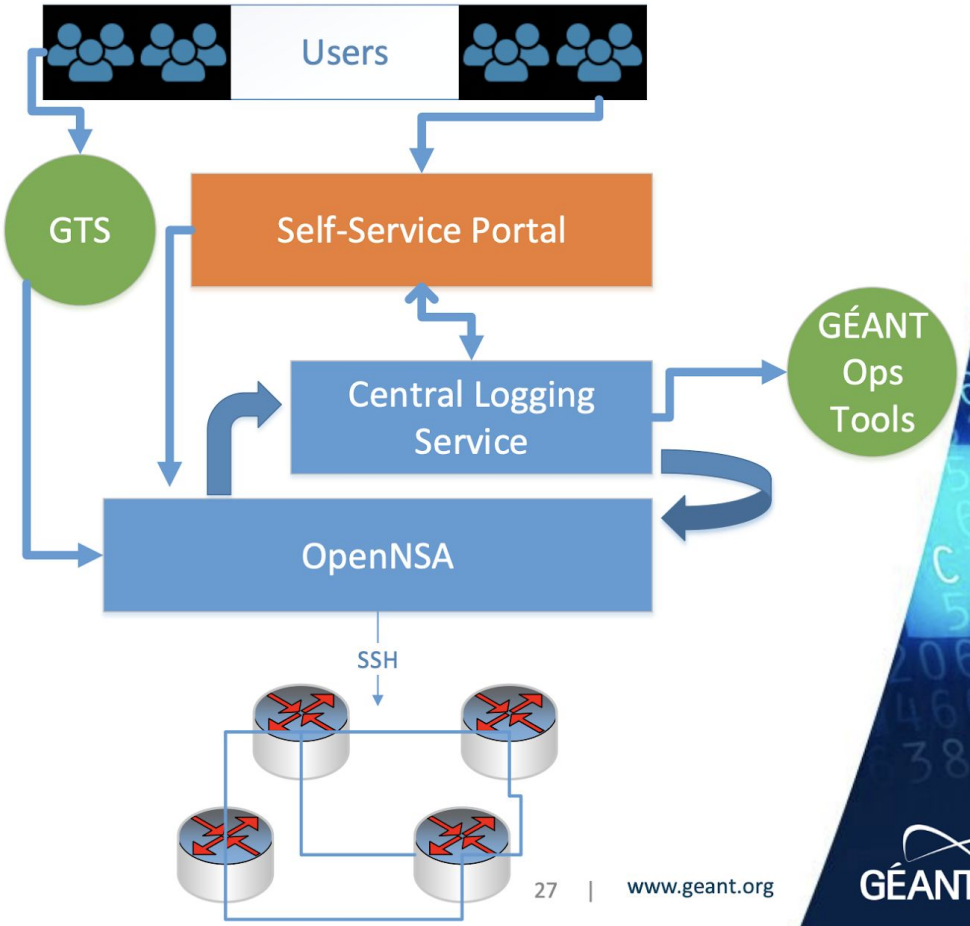
- ***Programmable*** – Software driven APIs to allocate core bandwidth as needed, and monitor status and performance.
- ***Scalable*** – Increased capacity scale and flexibility by leveraging latest technology (e.g. FlexGrid spectral partitioning, tunable wave modulation).
- ***Resilient*** – Protection and restoration functions using next generation Traffic Engineering (TE) protocols (e.g. Segment Routing (SR)).

Smart Services Edge

- ***Programmable*** – Software driven APIs to manage edge router/switch and retrieve telemetry information.
- ***Flexible*** - Data plane programmable switches (e.g. FPGA, NPU) in conjunction with compute resources to prototype new services (driven by Software Defined Networks (SDN))
- ***Dynamic*** – Dynamic instantiation of services using SDN paradigms (e.g. Network Function Virtualization (NFV, Virtual Network Functions (VNF), service chaining).

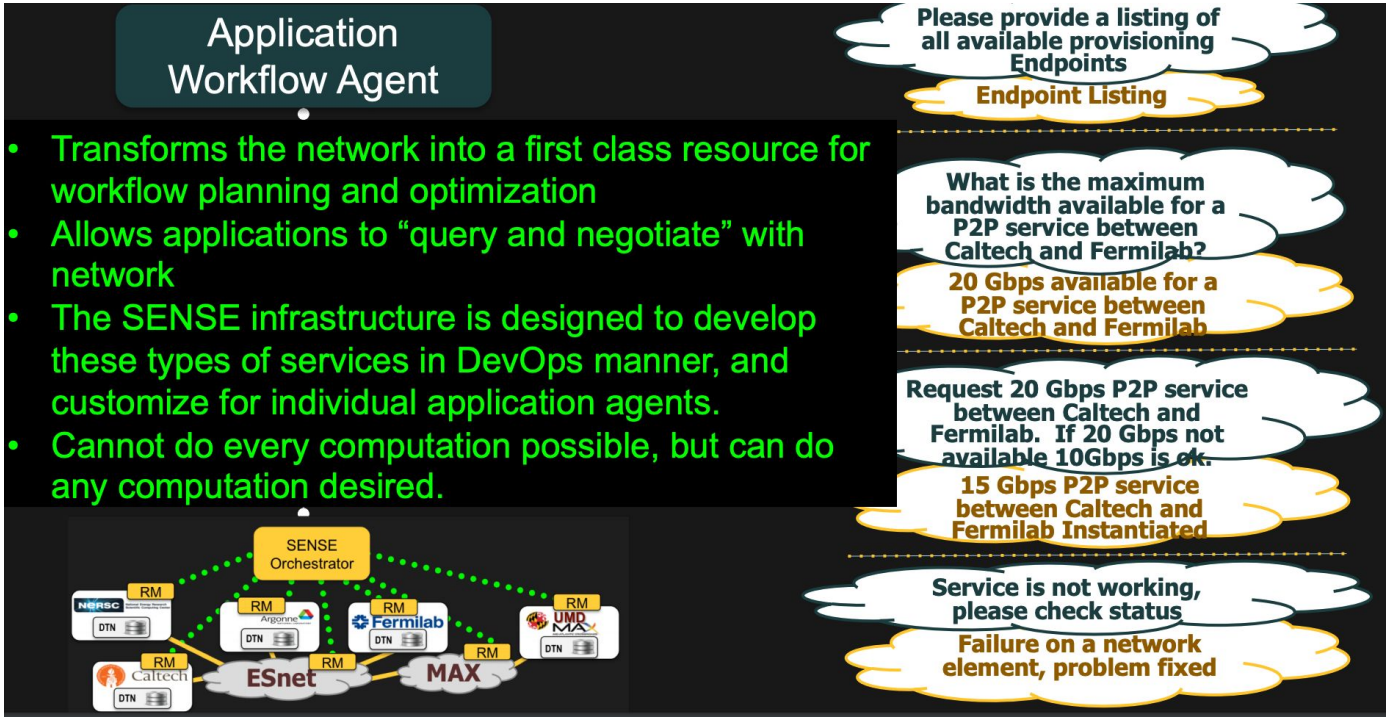
Connection Service

- Automated service delivery
- Self service portal
- Open interface (API) for use with 3rd party automated systems and orchestrators
- Automated monitoring and service inventory
- Multi-domain services



SENSE

SDN for End-to-end Networked Science at the Exascale (SENSE) - U.S. funded; ESNet, FNAL, ANL, NERSC, Caltech and Univ. of Maryland



Source: http://conferences.computer.org/scw/2018/#!/toc/3; https://indico.cern.ch/event/795430/

BigData Express

U.S DOE funded: FNAL, ESNet, StarLight, KISTI, Univ. of Maryland, ORNL

Our Solution - BigData Express

- BigData Express: a schedulable, predictable, and high-performance data transfer service

✓ – A peer-to-peer, scalable, and extensible data transfer model

– A visually appealing, easy-to-use web portal

✓ – A high-performance data transfer engine

– A time-constraint-based scheduler

✓ – On-demand provisioning of end-to-end network paths with guaranteed QoS

– Robust and flexible error handling

– CILogon-based security

Existing projects also in ATLAS (OVS btw AGLT2/MWT2/KIT), SDN aspects also in NSF-funded **SLATE**, **OSIRIS** and CERN's **NOTED** project

Source: http://conferences.computer.org/scw/2018/#!/toc/3; https://indico.cern.ch/event/777367/

Outlook

WG plans (schedule & meeting topics)

- Topics/meetings to be scheduled
- Programmable networks oriented:
  - GEANT survey and JISC/JANET evolution (Tim Chown, JISC)
  - GEANT connect (higher level services)
  - KIT network evolution, DFN evolution (Bruno)
- Cloud native oriented:
  - Network design @USC, Arista CloudVision (Azher Mughal)
  - Tungsten deployment update (CERN/Nikhef)
  - P4 and smartNIC technologies (Mauro Campanella, GARR)

Working Group Plans

- Finalise report by HEPiX fall
  - Possibility organise SDN/NFV session there
- Need volunteers to collect input on different areas
  - Cloud Native Networking (Cumulus Linux, ONOS, CNI, etc.)
  - SmartNICs
  - Programmable Networks (R&E plans)
- Identify potential areas for further work
  - Performance studies (incl. smartNICs)
  - Programmable Networks Prototypes
  - DCI testing/evaluation
  - WLCG DOMA activities
  - etc.

- Always looking for feedback and additional volunteers/sites for help.

Summary

- Explored several existing SDN/NFV approaches and use cases
- In Cloud Native Networking focusing primarily on open source approaches
  - CERN and Nikhef have active projects in deploying Tungsten Fabric
- SENSE and BigData Express leading projects in programmable networks and data transfers, but non-OpenFlow approaches are also being investigated (NOTED)
- Surveying R&E plans for higher-level services
- HEPiX will provide an opportunity to review status and provide Phase II input
- LHCONE/LHCOPN meeting at CERN in January 2020 should provide decision point on Phase II
- We welcome additional volunteers; contact us if you are interested!

References

WG meetings and notes: https://indico.cern.ch/category/10031/

SDN/NFV Tutorial: https://indico.cern.ch/event/715631/

Tungsten Fabric architectural overview:
https://tungstenfabric.github.io/website/Tungsten-Fabric-Architecture.html

OVN/OVS overview: https://www.openvswitch.org/

2018 IEEE/ACM Innovating the Network for Data-Intensive Science (INDIS) –
http://conferences.computer.org/scw/2018/#!/toc/3

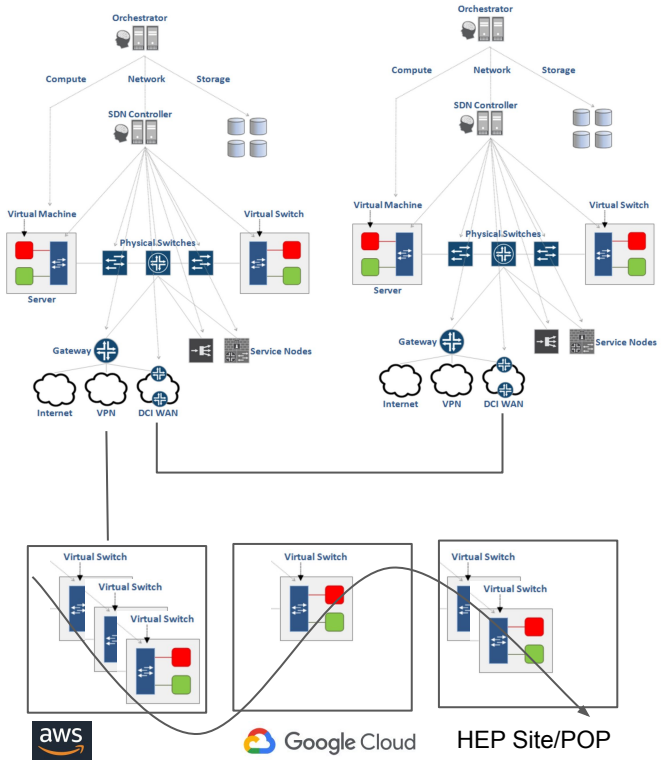
Cloud native Data Centre (book) -
https://www.amazon.com/Cloud-Native-Data-Center-Networking-Architecture/dp/1492045608/ref=sr_1_2?keywords=cloud+native+data+center+networking&qid=1568122189&s=gateway&sr=8-2

MPLS in the SDN Era (book) -
https://www.amazon.com/MPLS-SDN-Era-Interoperable-Scenarios/dp/149190545X/ref=sr_1_1?keywords=MPLS+in+the+SDN&qid=1568122219&s=gateway&sr=8-1

Backup slides

DC Edge - Multi-Cloud - DCI/Remote Compute

- SDN-based DC enables other interesting options
- **Data Center Interconnect (DCI)**
  - SDN services spanning multiple physical sites, each site with its own SDN deployment.
  - Agnostic to the Virtual Infrastructure Manager (Orchestrator) used.
- **Remote Compute**
  - Single SDN deployment extending its services to remote sites (POP/DC/Cloud). Ability to extend VPNs/VMs to another site without running a dedicated SDN cluster there.
- **Service chaining (NFV)**
  - Steering traffic between VPNs/VMs according to a policy, availability, etc.
- All the options are complementary and can be combined to create high-scale networking combining 100s or even 1000s of sites.



Networking Challenges

- Capacity/share for data intensive sciences
  - No issues wrt available technology, however
  - What if N more HEP-scale science domains start competing for the same resources ?
- Remote data access proliferating in the current DDM design
  - Promoted as a way to solve challenges within experiment's DDM
  - Different patterns of network usage emerging
    - Moving from large streams to a mix of large and small frequent event streams
- Integration of Commercial Clouds
  - Impact on funding, usage policies, security, etc.
- Technology evolution
  - Software Defined Networking (SDN)/Network Functions Virtualisation (NFV)

Technology Impact

- Increased importance to oversee network capacities
  - Past and anticipated network usage by the experiments, including details on future workflows
- New technologies will make it easier to transfer vast amounts of data
  - HEP quite likely no longer the only domain that will need high throughput
- Sharing the future capacity will require greater interaction with networks
  - While unclear on what technologies will become mainstream (see later), we know that software will play a major role in the networks of the future
  - We have an opportunity here
- It's already clear that software will play major role in networks in the mid-term
- Important to understand how we can design, test and develop systems that could enter existing production workflows
  - **While at the same time changing something as fundamental as the network that all sites and experiments rely upon**
  - We need to engage sites, experiments and (N)REN(s) in this effort

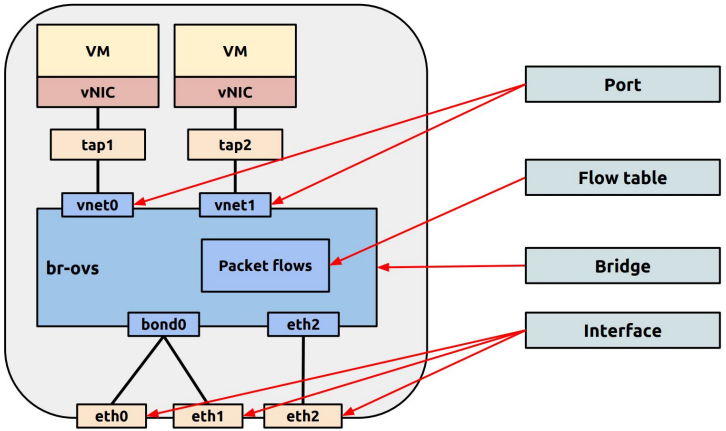
Software Defined Networks (SDN)

- Software Defined Networking (SDN) are a set of new technologies enabling the following use cases:
  - **Automated service delivery** - providing on-demand network services (bandwidth scheduling, dynamic VPN)
  - **Clouds/NFV** - agile service delivery on cloud infrastructures usually delivered via Network Functions Virtualisation (NFV) - underlays are usually Cloud Compute Technologies, i.e. OpenStack/Kubernetes/Docker
  - **Network Resource Optimisation (NRO)** - dynamically optimising the network based on its load and state. Optimising the network using near real-time traffic, topology and equipment. This is the core area for improving end-to-end transfers and provide potential backend technology for DataLakes
  - **Visibility and Control** - improve our insights into existing network and provide ways for smarter monitoring and control
- Many different point-to-point efforts and successes reported within LHCOPN/LHCONE
  - **Primary challenge is getting end-to-end!**
- While it's still unclear which technologies will become mainstream, it's already clear that software will play major role in networks in the mid-term
  - Massive network automation is possible - in production and at large-scale
- HEPiX SDN/NFV Working Group was formed to bring together sites, experiments, (N)RENs and engage them in testing, deploying and evaluating network virtualization technologies

Software Switches

Open vSwitch (OVS) - open source multilayer virtual switch supporting standard interfaces and protocols:

- OpenFlow, STP 802.1d, RSTP,
- Advanced Control, Forwarding, Tunneling
- Primarily motivated to enable VM-to-VM networking, but grew to become the core component in most of the existing open source cloud networking solutions



Runs as any other standard Linux app - user-level controller with kernel-level datapath including HW off-loading (recent) and acceleration (Intel DPDK)

Enables massive network automation …

Open vSwitch Features

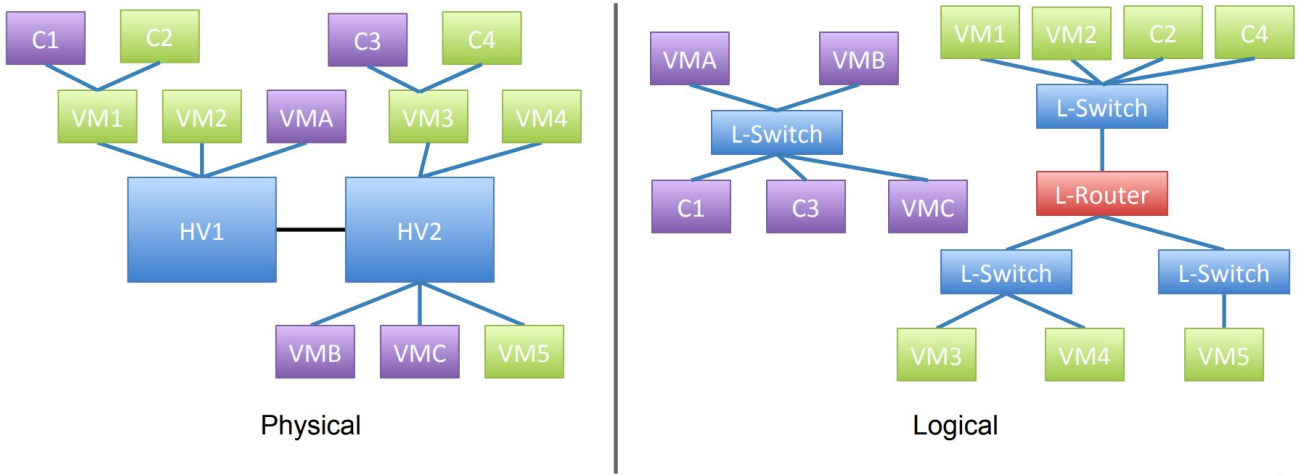
- Visibility into inter-VM communication via NetFlow, sFlow(R), IPFIX, SPAN, RSPAN, and GRE-tunneled mirrors
- LACP (IEEE 802.1AX-2008)
- Standard 802.1Q VLAN model with trunking
- Multicast snooping
- IETF Auto-Attach SPBM and rudimentary required LLDP support
- BFD and 802.1ag link monitoring
- STP (IEEE 802.1D-1998) and RSTP (IEEE 802.1D-2004)
- Fine-grained QoS control
- Support for HFSC qdisc
- Per VM interface traffic policing
- NIC bonding with source-MAC load balancing, active backup, and L4 hashing
- OpenFlow protocol support (including many extensions for virtualization)
- IPv6 support
- Multiple tunneling protocols (GRE, VXLAN, STT, and Geneve, with IPsec support)
- Remote configuration protocol with C and Python bindings
- Kernel and user-space forwarding engine options
- Multi-table forwarding pipeline with flow-caching engine
- Forwarding layer abstraction to ease porting to new software and hardware platforms

Controllers - Open DayLight

- Modular open platform for customizing and automating networks of any size and scale. Core use cases include:
  - **Cloud and NFV** - service delivery on cloud infrastructure in either the enterprise or service provider environment
  - **Network Resource Optimisation** - Dynamically optimizing the network based on load and state; support for variety of southbound protocols (OpenFlow, OVSDB, NETCONF, BGP-LS)
  - Automated Service Delivery - Providing on-demand services that may be controlled by the end user or the service provider, e.g. on-demand bandwidth scheduling, dynamic VPN
  - Visibility and Control - Centralized administration of the network and/or multiple controllers.
- Core component in number of open networking frameworks
  - ONAP, OPNFV, OpenStack, etc.
- Integrated or embedded in more than 50 vendor solutions and apps
- ODL is just one of many controllers that are available:
  - OpenContrail, ONOS, MidoNet, Ryu, etc.

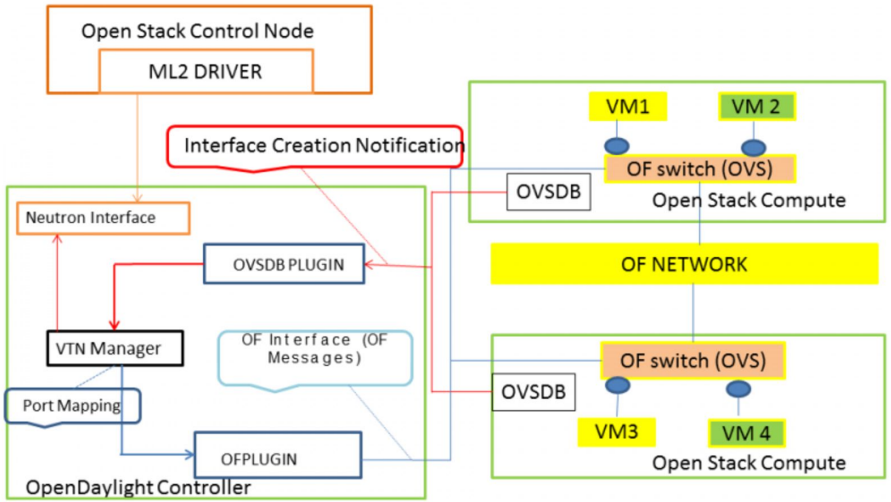
Controllers - Open Virtual Network (OVN)

- Open source logical networking for OVS
- Provides L2/L3 networking
  - Logical Switches; L2/L3/L4 ACLs
  - Logical Routers, Security Groups
  - Multiple Tunnel overlays (Geneve, VXLAN)
  - Top-of-rack-based & software-based physical-to-logical gateways



Cloud Compute - OpenStack Networking

- Cloud stresses networks like never before
  - Massive scale, Multi-tenancy/high density, VM mobility
- OpenStack Neutron offers a plugin technology to enable different (SDN) networking approaches - brings all previously mentioned techs together



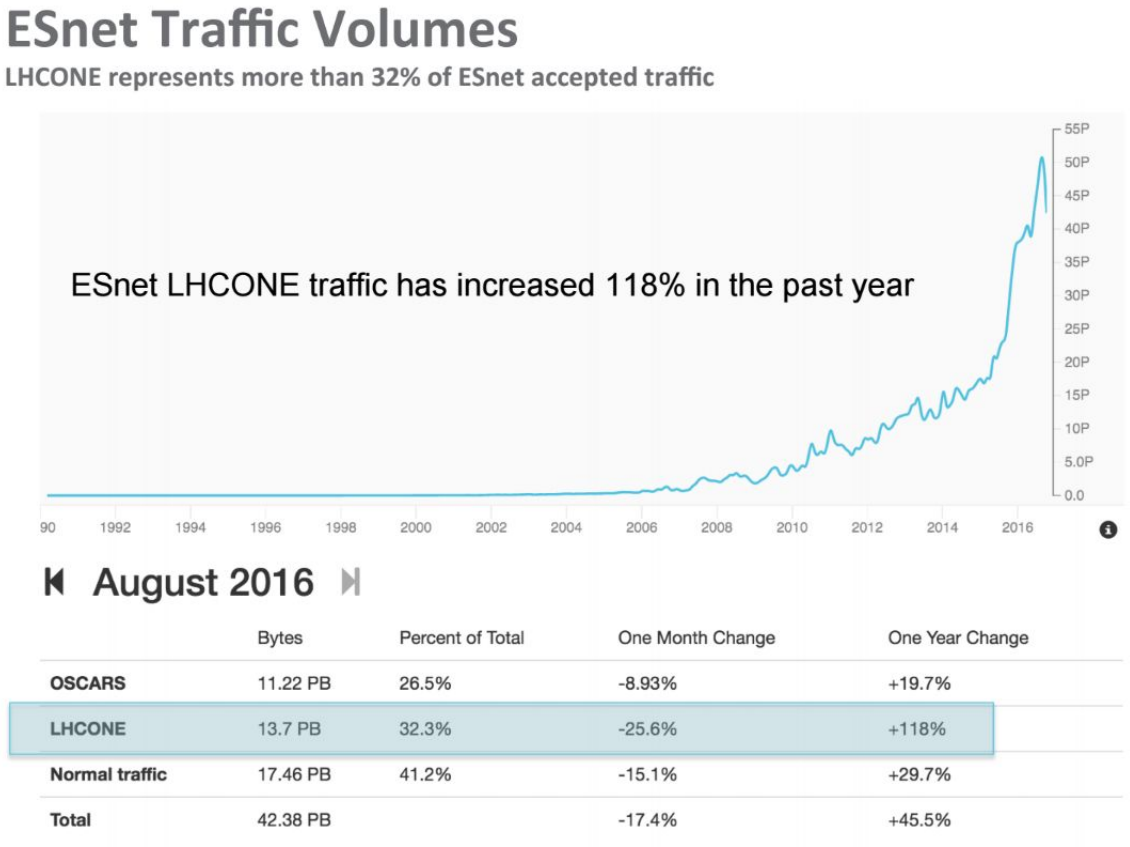
ML2 driver is what makes controllers pluggable, so you can easily replace Neutron controller with OpenDaylight, OVN, etc.

Both generic and vendor-specific plugins are available

Cumulus Linux

- Alternative to OVS - uses separate apps/kernel functions to program different functionality such as STP/RSTP (mstpd), VXLAN (ifupdown2), VLAN (native linux bridge) etc.
- It does contain OVS to enable integration with controllers:
  - VMware NSX, Midokura Midonet, etc.
- Unlike OVS, Cumulus Linux is not an app, but a distribution, which is certified to run on bare metal switches
  - The list of supported HW is at (https://cumulusnetworks.com/products/hardware-compatibility-list/)
  - Mainly Broadcom Tomahawk, Trident2/+, Helix4 and Mellanox Spectrum ASICs
- Otherwise runs like standard Linux, which means compute and network “speak the same language”
  - E.g. automation with Ansible, Puppet, Chef, etc.

R&E Traffic Growth Last Year



	Bytes	Percent of Total	One Month Change	One Year Change
OSCARS	11.22 PB	26.5%	-8.93%	+19.7%
LHCONE	13.7 PB	32.3%	-25.6%	+118%
Normal traffic	17.46 PB	41.2%	-15.1%	+29.7%
Total	42.38 PB		-17.4%	+45.5%

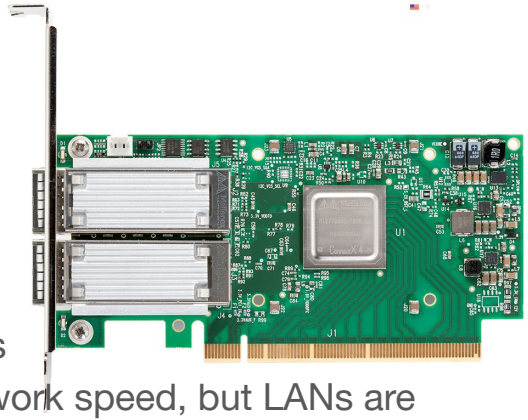
Slide from Michael O'Connor, LHCONE operations update

In general, ESNet sees overall traffic grow at factor 10 every 4 years. Recent LHC traffic appears to match this trend.

GEANT reported LHCONE peaks of over 100Gbps with traffic increase of 65% in the last year.

This has caused stresses on the available network capacity due to the LHC performing better than expected, but the situation is unlikely to improve in the long-term.

WAN vs LAN capacity



- Historically WAN capacity has not always had a stable relationship compared to data-centre
  - In recent history WAN technologies grew rapidly and for a while outpaced LAN or even local computing bus capacities
  - Today 100Gbps WAN links are the typical high-performance network speed, but LANs are also getting in the same range
    - List price for 100Gbit dual port card is ~ $1000, but significant discounts can be found (as low as $400), list price for 16 port 100Gbit switch is $9000
- Today it is easy to over-subscribe WAN links
  - in terms of $ of local hardware at many sites
- Will WAN be able to keep up ? **Likely yes,** however:
  - We did benefit from the fact that 100Gbit was deployed on time for Run2, might not be the case for Run3 and 4
  - By 2020 800 Gbps waves likely available, but at significant cost since those can be only deployed at proportionally shorter distances
- Planning of the capacities and upgrades (NREN vs sites) will be needed

Improving Our Use of the Network

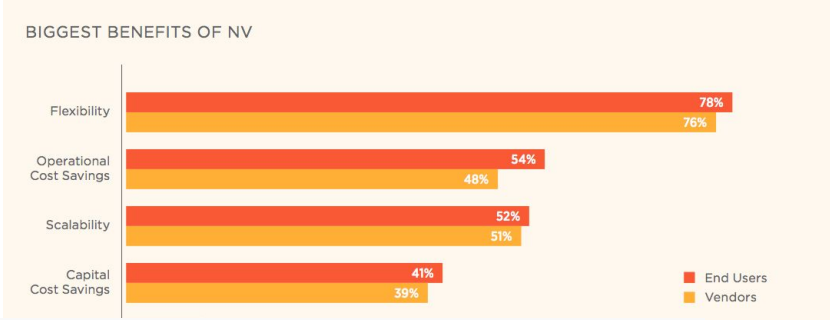
- TCP more stable in CC7, throughput ramp ups much quicker
  - Detailed report available from Brian Tierney/ESNet
- Fair Queueing Scheduler (FQ) available from kernel 3.11+
  - Even more stable, works better with small buffers
  - Pacing and shaping of traffic reliably to 32Gbps
- Best single flow tests show TCP LAN at 79Gbps, WAN (RTT 92ms) at 49Gbps
  - IPv6 slightly faster on the WAN, slightly slower on the LAN
- **In summary: new enhancements make tuning easier in general**
  - But some previous “tricks” no longer apply
- New TCP congestion algorithm (TCP BBR) from Google
  - Google reports factor 2-4 performance improvement on path with 1% loss (100ms RTT)
  - Early testing from ESNet less conclusive and questions need answering

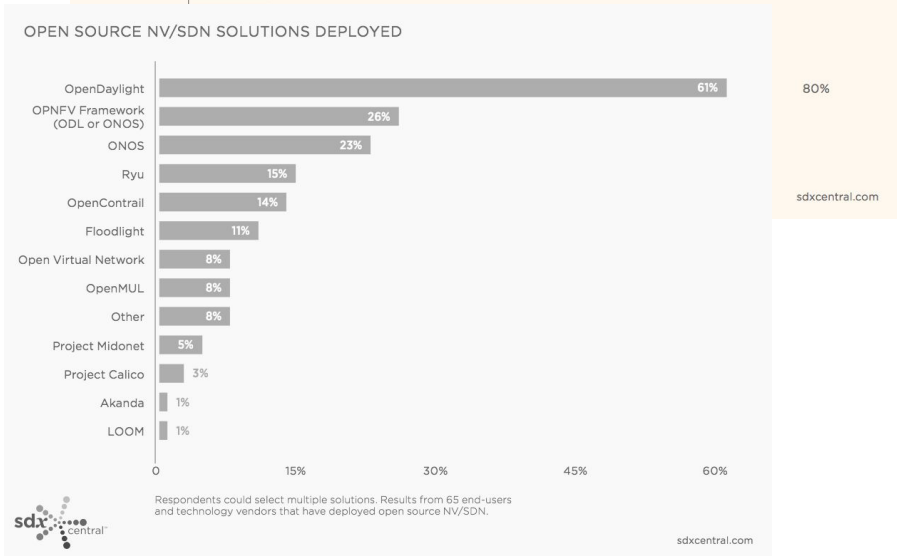
R&E Networking

- R&E network providers have long been working closely with HEP community
  - HEP has been representative of the future data intensive science domains
  - Often serving as testbed environment for early prototypes
- Big data analytics requiring high throughput no longer limited to HEP
  - SKA (Square Kilometer Array) plans to operate at data volumes 200x current LHC scale
  - Besides Astronomy there are MANY science domains anticipating data scales beyond LHC, cf. ESRFI 2016 roadmap
- **What if N more HEP-scale science domains start competing for the same network resources ?**
  - Will HEP continue to enjoy “unlimited” bandwidth and prioritised attention or will we need to compete for the networks with other data intensive science domains ?
  - Will there be **AstroONE**, **BioONE**, etc., soon ?

Tech Trends: Software Defined Networks (SDN)

- SDN is a set of technologies offering solutions for many of the future challenges
  - Current links can handle ~ 6x more traffic if we could avoid peaks and be more efficient
  - SDN driven by commercial efforts
- Many different point-to-point efforts and successes reported within LHCOPN/LHCONE
  - **Primary challenge is getting end-to-end!**
- While it's still unclear which technologies will become mainstream, it's already clear that software will play major role in networks in the mid-term
  - Will experiments have effort to engage in the existing SDN testbeds to determine what impact it will have on their data management and operations ?





Tech Trends: SD-WAN

- Large Network as a Service providers include several well established CSPs such as Amazon, Rackspace, AT&T, Telefonica, etc.
- Recently more niche NaaS providers have appeared offering SD-WAN solutions
  - Aryaka, Cloudgenix, Pertino, VeloCloud, etc.
  - Their offering is currently limited and not suitable for high throughput, but evolving fast
- SD-WAN market is estimated to grow to $6 billion in 2020 (sdxcentral)
- Will low cost WAN become available in a similar manner we are now buying cloud compute and storage services ?
  - Unlikely, our networks are shared, not easy to separate just LHC traffic
  - Transit within major cloud providers such as Amazon currently not possible and unlikely in the future, limited by regional business model - but great opportunity for NRENs

Tech Trends: Containers

- Recently there has been a strong interest in the container-based systems such as Docker
  - They offer a way to deploy and run distributed applications
  - Containers are lightweight - many of them can run on a single VM or physical host with shared OS
  - Greater portability since application is written to container interface not OS
- Obviously networking is a major limitation to containerization
  - Network virtualization, network programmability and separation between data and control plane are essential
  - Tools such as Flocker or Rancher can be used to create virtual overlay networks to connect containers across hosts and over larger networks (data centers, WAN)
- Containers have great potential to become disruptive in accelerating **SDN** and **merging LAN and WAN**
  - But clearly campus SDNs and WAN SDNs will evolve at different pace

Network Operations

- Deployment of perfSONARs at all WLCG sites made it possible for us to see and debug end-to-end network problems
  - OSG is gathering global perfSONAR data and making it available to WLCG and others
- A group focusing on helping sites and experiments with network issues using perfSONAR was formed - WLCG Network Throughput
  - Reports of non-performing links are actually quite common (almost on a weekly basis)
  - Most of the end-to-end issues are due to faulty switches or mis-configurations at sites
  - Some cases also due to link saturation (recently in LHCOPN) or issues at NRENs
- Recent network analytics of LHCOPN/LHCONE perfSONAR data also point out some very interesting facts:
  - Packet loss greater than 2% for a period of 3 hours on almost 5% of all LHCONE links
- Network telemetry (real-time network link usage) likely to become available in the mid-term (but likely not from all NRENs at the same time)
- It is increasingly important to focus on site-based network operations