Networking - Best Practice for Sites, and Evolution

Shawn McKee, Marian Babik

ATLAS Sites Jamboree

CERN, March 8, 2019

Geneva, Switzerland

Introduction

- During this ATLAS Sites Jamboree we have covered many topics. I want to provide **some planning and best practice suggestions** for site networking
  - Networking use will increase as we move toward Run-3 and HL-LHC
- There are four main areas to cover
  - Capacity
  - Capability
  - Monitoring and Security
  - Outlook
- First a little overview of where we are and where we are going

Current Network Status

This talk will focus on the mid-to-long term planning for ATLAS sites but I did want to state a few things about our current status up front

- Our networks have performed very well for our community
- Most physicists are happy with the networking we have
- Primary concerns exist around our ability to fully utilize **existing** networks
- Visibility is key to understanding, maintaining and fixing our networks

So there continues to be **near-term** work regarding our networking in optimizing, monitoring and fixing network problems, **but** we should also think longer term regarding how the situation may evolve and what that might mean for us.

Context, Considerations

- HEP network data transfer volume continues to rise at roughly a factor of 10 every 4 years
- We are now in LS2 and have an opportunity to develop our software and hardware infrastructures to be ready for Run-3 and, eventually, HL-LHC
- Our **ATLAS** sites need to plan for how, and on what timescale, they will evolve their networking infrastructure.
- What about **Clouds** and **HPCs**? What role will site network play to enable us to effectively utilize these kinds of resource?

Capacity

Capacity

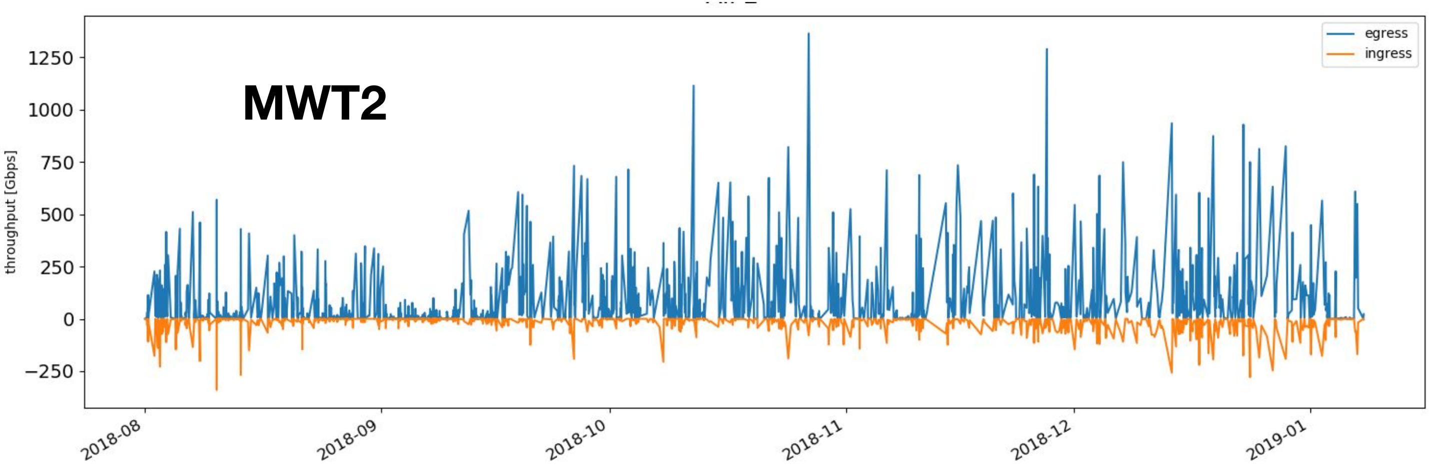
- When discussing networking, the first parameter usually discussed is the bandwidth (capacity), either in the LAN or to the WAN
- Sites usually have a pretty good handle on the LAN capacity required to allow their compute and storage to operate smoothly, but the WAN capacity is less well understood
  - For **Run-1** we targeted **1-10G** for a typical Tier-2 center while larger sites and Tier-1s should have **2-4x10G**
  - For **Run-2** we targeted **10G as a minimum** for Tier-2s with a **goal** of **multiple 10G** or **40G** as desirable. **Large sites** (Tier-1 and Tier-2) were encouraged to have **multiple 40G** to **100G** connectivity
  - Report by Bos-Fisk: https://twiki.cern.ch/twiki/bin/view/LHCOPN/T2sConn

Network Capacity Estimation for Run3

- Assume analysis on an HT-Core (job-slot) consumes 1.2 MBytes/sec
  - Implies job-slots need that level of network bandwidth to storage
  - WAN access to remote storage at 20% (ATLAS avg now)
    - 10 PBytes/day, 8 PB LAN, 2 PB WAN from Mario
  - **Minimal Tier-2**: 1000 job slots => 1.2 GBytes/sec, WAN 1.6 Gbits/sec
  - **Nominal Tier-2**: 5000 job slots => 6 GBytes/sec, WAN 9.6 Gbits/sec
  - **Leadership Tier-2**: 10000 job slots => 12 GBytes/sec, WAN 19.2 Gbits/sec
  - **NOTE**: Run-3 will have 3-4 times the data...have to either increase cores or improve average software throughput by that factor
- **Summary Network Capacity Recommendation:**
  - Average numbers above need a burst capability, assume x3
  - **Minimal Tier-2 WAN**: 1.6 Gbps x 3 = 4.8 Gbps => **10G link**
  - **Nominal Tier-2 WAN**: 9.6 Gbps x 3 = 28.8 Gbps => **40G link**
  - **Leadership Tier-2 WAN**: 19.2 Gbps x 3 = 57.6 Gbps => **80G link**

Capacity from XCache Use-Case

From Ilija's presentation yesterday we see the inferred LAN (blue) and WAN (orange)) traffic from his XCache simulation. Assumes **40 MBytes/s** per job which is much larger than ATLAS avg **300KB/sec**



Comparing with the numbers from the previously slide. The WAN/LAN peaks are a factor of **33 times** larger because of the 40 MBytes/sec vs **1.2 MBytes/sec** per job. Top LAN value from from 1.35 Tbps down to **41 Gbps**, similarly WAN peak of **300 Gbps** drops to **9.5 Gbps**.

Perhaps XCache could save some WAN bandwidth. The above neglects the origin requirements.

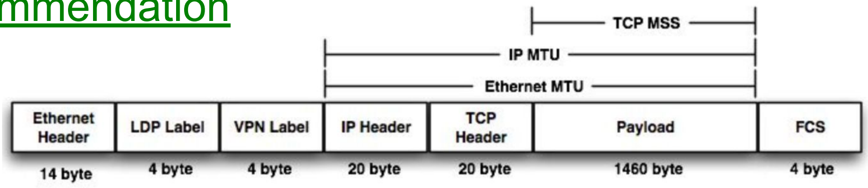
Impact of latency and loss

- A number of factors impact our ability to **use the nominal capacity**
- Latency/delay - impacts TCP buffers, easily calculated via Bandwidth-delay product (BDP)
  - tcp buffer to reach 10 Gbps with RTT of 30.0 ms >= 36.62 MBytes
  - tcp buffer to reach 10 Gbps with RTT of 60.0 ms >= 73.242 MBytes
  - tcp buffer to reach 10 Gbps with RTT of 250.0 ms >= 305.17 MBytes
- This relates not only to TCP host tuning on DTNs, but also to network equipment selection
- Vendors rarely put information on packet buffers in their data sheets
  - TOR switches (if cheap) **won't have enough buffers** to transfer over higher latencies
  - Stateful firewalls often have very small buffers
  - At the same time large buffers increase latency (bufferbloat),

- Consistent packet-loss can have huge impact (BDP/loss calculator)
  - network limit (MSS 1460 byte, RTT: 30.0 ms, Loss: 10^{-08}): 3893.33 Mbit/sec
  - network limit (MSS 1460 byte, RTT: 200.0 ms, Loss: 10^{-08}) : 584.00 Mbit/sec.
- Mixing UDP and TCP transfers at high throughput will likely cause TCP to suffer

Impact of MTU, protocols and routing

- MTU is one of the top operational issues impacting performance/connectivity
  - LHCOPN/LHCONE is working on a recommendation
  - MTU issues combined with load-balancing are very challenging to debug



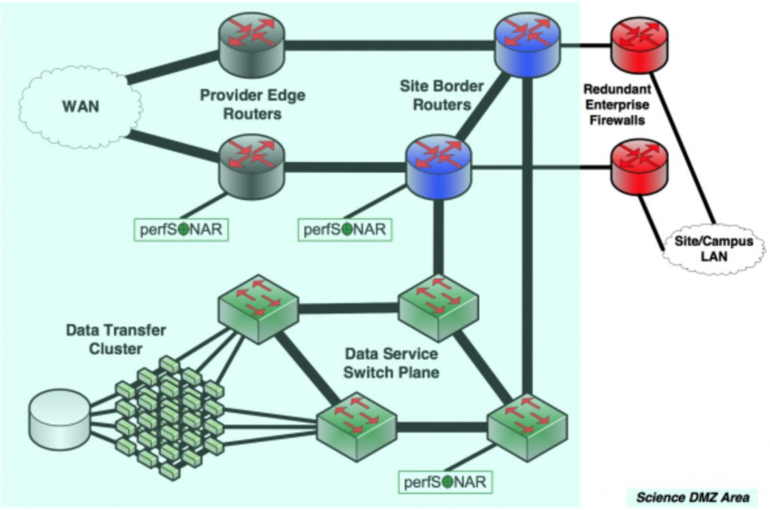
- IPv4 and IPv6 performance could be very different
  - IPv6 is likely to be processed by different branches of code (QoS, firewalls, IPv6 TCP stack, etc.) or even different equipment
  - Check network path first before looking any further
  - Establishing expectations and test them for both IPv4 and IPv6 is important

- Network paths are dynamic and while sites usually have limited control over this, change of route can have major impact on capacity. This applies also to Commercial Clouds (which will likely take commercial routes unless you have direct peering)

Capability

Network Design

- Core capabilities to effectively deploy and support high-performance science applications include high bandwidth, advanced features, and capable gear that does not compromise on performance
- Science DMZ architecture is one of the examples that brings together performance, operational and security requirements
- Concerning network components:
  - Make sure routers/switches has sufficient buffer space to handle “fan-in” issues
  - Be wary of routers and switches that are over-subscribed (as this leads to consistent loss)
  - Look for devices that have flexible and performant ACLs support to eliminate need for stateful firewall which impact performance
  - If you’re planning to re-engineer your network, consider SDN/NFV approaches



WAN Connectivity: R&E Networks

- For typical T2 connecting to LHCONE is likely the best option
  - LHCONE use avoids performance killing security devices and provides better traffic visibility to more quickly identify and fix network issues when they occur
- However, LHCONE is a federation of networks (not a single monolithic network) - BGP issues can pop up from time to time and impact both reachability and performance
- Network paths in LHCONE are determined based on peering and transit agreements between R&Es (budget)
  - Not creating additional capacity
- Inefficient use of network causing congestions pressures R&Es (and sites!) to upgrade (trigger is variable, sometimes as soon as 20% average capacity is reached)
- BGP Routing issues can be explored using existing tools:
  - LHCONE looking glass **prototype** - http://lhcone-lg.cern.ch
  - GEANT looking glass - https://tools.geant.net/portal/links/lg/
  - ESNet - https://my.es.net/

Network Optimisation

- 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. See http://fasterdata.es.net/
- New TCP congestion algorithm (TCP BBR) from Google
  - Google reports factor 2-4 performance improvement on path with 1% loss (100ms RTT)
  - Experimental evaluation by KIT is less conclusive
  - However this is a work in progress, BBR version 2 is in the works (not yet open source)
  - It will likely become serious contender and we’ll need to plan its evaluation/deployment
- There are TCP alternatives to explore, which can completely change all this

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 performance using perfSONAR - WLCG Network Throughput
  - Reports of non-performing links are actually quite common
- Sites experiencing **known** network issues should first contact their local network team, who can pursue the issues with the regional and backbone providers on their behalf
- **LHCONE operations** - support for establishing and operating LHCONE infrastructure - regular meetings and support mailing list
- LHCOPN/LHCONE workshop - organised bi-annually - good place to meet R&Es and discuss architecture and plans
  - Next one is in Umea, Sweden, https://indico.cern.ch/event/772031/)

Cloud Networking Considerations

- If you're planning to re-engineer your network stack and/or planning to deploy **OpenStack** or **Kubernetes** consider Software Defined Networking (SDN) approaches
- Orchestrating network together with Compute is possible and can work in production and at scale. When selecting an approach consider the following aspects:
  - **Multistack** - Connecting multiple orchestration stacks like Kubernetes, Mesos/SMACK, OpenShift, OpenStack and VMware
  - Networking and security across legacy, virtualized and containerized applications
  - Multistack and across-stack policy control, visibility and analytics
  - Multi-cloud support - DCI and Remote Compute
  - **Support for configuration and control** of the network equipment
  - Offloading of virtual networks via physical hardware (or via smart NICs)
- Some of the **key benefits**: self-service networks, auto-provisioning of VPNs, isolation (multi-domain), improved visibility and debugging of the networks, scalability (spanning services across multiple data centers), etc.

- 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

Monitoring and Security

Network Monitoring Considerations

- We need network visibility to understand performance, find problems and enable orchestration
- **All sites should have deployed perfSONAR and have a plan to keep the hardware and software updated**
    - The recommendation is to provide two instances: latency and throughput (which could be on the same server with at least two NICs)
    - The perfSONAR instances should be (co)located with your sites STORAGE, network-wise
    - The throughput instance should use the same NIC capacity as your storage servers
    - Additional perfSONAR instances can be helpful for identifying LAN issues
- https://opensciencegrid.org/networking/perfsonar/installation/#perfsonar-installation-guide

Security Considerations

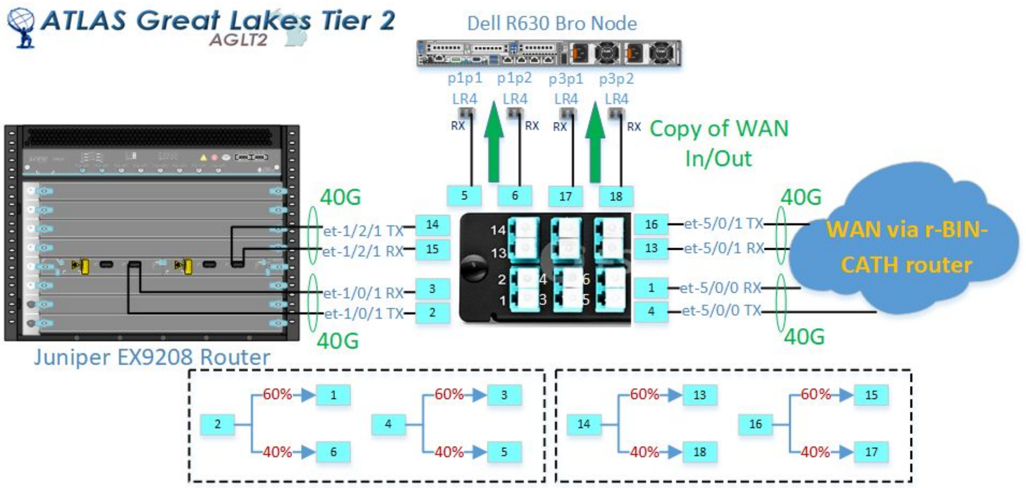
Enabling some additional security via traffic monitoring can help your site identify and defend against attacks

The **WLCG Security Operations Center** working group is advocating a solution involving **Zeek** (formerly "Bro"), **MISP** and **Elastiflow**

Info at http://wlcg-soc-wg-doc.web.cern.ch/wlcg-soc-wg-doc/

I have deployed an optical tap and Zeek monitoring at the AGLT2 UM site for less than **$2K** and the use of a worker node.

As part of site planning it would be useful to consider deploying a similar capability as the network is upgraded.



Outlook

Outlook

- Long term outlook (5-10 years) will likely involve:
  - Capacity sharing - other big research domains coming online
  - (Re)organisation - evolution of LHCONE (ASTRONE ?), possibly some form of SD-WAN (dynamic circuits/L3 VPNs on demand)
  - Cloud networking - network federations spanning multiple data centres (inc. commercial clouds), ability to develop and operate services across large number of heterogenous sites easily from one location
- Ready ?

Conclusion

All ATLAS sites should be reviewing their current network setup and planning to make needed changes in time for Run-3

The recommendations provided are based upon the current computing model.

- Changes to that could easily have implications on the network recommendations

Many details concerning design of the LAN are strongly dependent upon local site details and are hard to provide detailed recommendations for.

Questions?

Additional Slides

Cloud Networking

- Commercial cloud providers already operate big networks at global scale with significantly higher capacities that are available in R&E
- Cloud computing is also becoming an important topic and eventually we'll need to find ways how to effectively bridge commercial and R&E networks
  - ATLAS/Rucio project with Google is one example going in this direction

- Will commercial WAN become available in a similar manner we are now buying cloud compute and storage services ?
  - The underlying cost will be decisive
  - Transit within major cloud providers such as Amazon/Google currently not possible and likely challenging in the future, limited by regional business model - but great opportunity for NRENs

Software Defined Networks (SDN)

- Software Defined Networking (SDN) 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

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

Data Lakes

- Simone and Rob have already talked on this, but a few comments relevant to networking:
- Data Lakes will rely on the networks to provide both in-lake and out-of-lake transport
  - Decoupling storage and data management is an opportunity for us to re-think how we currently manage and operate our networks
  - Finding ways how to integrate network monitoring/feedback with storage and provide near-real time information to the controllers will be important to get an efficient system
- Streaming/Caching is another area which is currently unaware of the network
  - Non-managed caches directly in the network hubs could provide significant benefits
  - This is already provided by commercial CDNs and is a potential opportunity for (N)RENs
  - It could be coupled with other services CDNs provide such as security (DoS protection, etc)
- Unless we find ways how to better interact with the networks, network operational cost of data lakes can be significant
  - Debugging ASGC to NDGF transfers where storage is located in Slovenia gives a hint at the challenges we'll likely face

R&E Networking - provisioning

- 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
- Other data intensive sciences will be coming online
  - 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
- Network provisioning will need to evolve
  - LHCOPN as a dedicated/private network might become shared with other experiments
    - Already the case for Belle II, other experiments might be coming
    - Understanding how we monitor/share existing capacities will be needed
  - LHCONE has its own challenges - will we see AstroONE, BioONE ?
    - In its present status - it's already extensively complex
- **Monitoring and managing network as a resource in a similar way we do compute and storage today is becoming likely in the future**

Summary

- Increased importance to oversee network capacities
  - Past and anticipated network usage by the experiments, including details on future workflows
  - Sites vs NREN capacities
- New technologies will make it easier to transfer vast amounts of data
  - And HEP 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, we know that software will play a major role in the networks of the future
  - We have an opportunity here with a little planning
- Containers might become the “accelerator” for adoption of SDNs on campus
  - With impact on skills and effort required to manage local networks

Questions or Comments?

LHC schedule

We will see significant pressure on network resources, which will likely accelerate in HL-LHC (x10). Major increases in funding are not expected and will likely remain flat.