Snowmass CompF4 Topical Group Workshop

Networking

Eli Dart (ESnet)
Chin Guok (ESnet)
Shawn McKee (UMich)

Snowmass / Seattle
July 19, 2022

Sources of Input

- ESnet HEP Requirements Review Report
  - Jason Zurawski, Ben Brown, Dale Carder, Eric Colby, Eli Dart, Ken Miller, Abid Patwa, Kate Robinson, Lauren Rotman, Andrew Wiedlea, “2020 High Energy Physics Network Requirements Review Final Report”, ESnet Network Requirements Review, June 29, 2021, LBNL LBNL-2001398
- Snowmass CompF4 Whitepapers
  - Alex Sim, Ezra Kissel, Chin Guok. ”Deploying in-network caches in support of distributed scientific data sharing“, arXiv:2203.06843 [cs.NI] (pdf).
  - Tom Lehman, Xi Yang, Chin Guok, Frank Wuerthwein, Igor Sfiligoi, et al. ”Data Transfer and Network Services Management for Domain Science Workflows”, arXiv:2203.08280 [cs.NI] (pdf).
  - Yatish Kumar, Stacey Sheldon, Dale Carder, “Transport Layer Networking”, arXiv:2203.02861 [cs.NI] (pdf).
- Engagements/Collaborations
  - Research Networking Technical Working Group (RNTWG)
  - WLCG Data Challenges (2021 - 2028) - Lessons Learned
  - Global Network Advancement Group (GNA-G)

Framing and Context

- Networking is a central component of CompF4 (storage and processing resource access).
  - Networking is the data circulatory system for scientific collaborations, transporting science data (the “crown jewels” of the science community) to computing and data analysis, and the results back to the collaboration
- Technology changes quickly; people change slowly; nature does not change
  - It is critical that Physics be able to effectively use technology to make progress
  - Networking is a critical foundation and a key enabler of computing and data analysis for Physics (and of course for other disciplines too!)
  - Physics needs to have the people to effectively integrate networking into larger Physics processes - whatever the technology stack of the day might be
- We describe four sub-topics that frame a strategic path forward to ensure that Physics is able to make optimal use of networking into the future

Networking Section Outline

- Networking section has 4 sub-topics
  - **Network Interaction Optimization** - Capabilities or functions that allow the application to better interact with the network, resulting in improved performance or enhanced features.
  - **Resource Orchestration and Automation** - The ability to coordinate the scheduling and provisioning of network resources to facilitate predictable data movement behaviors.
  - **Network and Traffic Visibility** - Insight into network health and traffic flow patterns to guide data movement decisions, and direct troubleshooting efforts.
  - **Data Movement Optimization** - Capabilities or functions that can improve the end user experience by reducing the time to fetch data.
- The capabilities outlined in the 4 sub-topics are discrete but interrelated.
- Integration of these capabilities will result in a whole that is greater than the sum of its parts.

Sub-Topic 1 - Network Interaction Optimization

- **What is it?**
  - Traffic shaping and pacing, IPv6, and source-based routing.
- **Why?**
  - Optimizing data movement is not a single big issue, but many small enhancements. Traffic shaping and pacing can reduce congestion in the network by preventing bandwidth oversubscription, source-based routing can provide path engineering mechanisms to facilitate more predictable transfer behaviors, and IPv6 can be leveraged to address IPv4 address exhaustion (along with compliance with the OMB IPv6-only mandate).
- **Examples of current efforts**
  - Ingress network shaping, e.g., to reduce packet drops due to interface speed mismatches.
  - Intelligent pacing, e.g., BBRv2 congestion control
  - IPv6 (only) transition, e.g., OMB IPv6 IPT docs, CERN IPv6-only deployments
  - (Multi-domain) source based routing, e.g., SRv6
- **Considerations**
  - Data movement performance is an end-to-end capability. This requires all participating domains to offer congruent services and capabilities - in production as well as in development.
  - Coordinated actions from stakeholders are necessary for this effort to be successful and maintainable.

Sub-Topic 2 - Resource Orchestration and Automation

- **What is it?**
  - Application driven intelligent orchestration and automation of DTNs, and network resources.
- **Why?**
  - By orchestrating workflow dependent resources and receiving the appropriate commitments, Service Level Expectations (SLEs) or Service Level Agreements (SLAs) can be established to provide predictable behavior.
  - SDN provides the control-plane abstractions and portability needed to implement the desired data-plane behaviors.
  - Simplistic allocation of resources for any given request can result in sub-optimal commitment of overall resources and lead to resource fragmentation. Leveraging AI/ML to predict usage and help drive resource allocation decisions can help provide more optimal solutions.
- **Examples of current activities.**
  - Site traffic steering, e.g., NOTED
  - WAN and DTN orchestration, e.g., SENSE, SENSE/Rucio collaboration
  - HPC API, e.g., NERSC Superfacility API
  - Whitebox switches and next-generation routing, e.g., RARE
  - AI/ML driven network utilization prediction and traffic engineering, e.g., HECATE
  - Integrated facilities, e.g., DOE IRI ABA
- **Considerations**
  - Standardization of orchestration API would facilitate scaling of such services in the larger ecosystem, vs bespoke APIs for each resource.
  - A common AuthN method would simplify access to these services.
  - Verifiable AI is essential to build trust needed to hand over control of the network.
  - Troubleshooting complexity increases if the network is constantly adapting.
  - “Break-glass” processes should be implemented in the event that the AI is compromised.

Sub-Topic 3 - Network and Traffic Visibility

- **What is it?**
  - Precision network telemetry and high-fidelity traffic flow tracking
- **Why?**
  - Precision network telemetry information that is accessible to applications can be extremely valuable when making intelligent decisions on when data movements should be scheduled. This can help in setting expectations, as well as understanding performance issues.
  - High fidelity traffic flow tracking is important in accurate data movement analysis and auditing, in addition to developing usage models.
  - We have great visibility on the computing side - we need better visibility on the networking side
- **Examples of current activities**
  - RNTWG packet marking and flow labeling, e.g., Firefly packet effort
  - ESnet High-Touch precision network telemetry services
  - P4 In-Network Telemetry, e.g., Q-Factor
- **Considerations**
  - A unified statistics platform across WLCG sites would go far to facilitate end-to-end multi-domain traffic analysis.
  - A common AuthN framework would be beneficial if sensitive data needs to be accessed.

Sub-Topic 4 - Data Movement Optimization

- **What is it?**
  - In-network caching, multi-path end-to-end load-balancing, meta-scheduling.
- **Why?**
  - In-network caching can reduce the time to retrieve data, improving workflow performance. This is especially true if the placement of the data is geographically local. An added benefit to in-network caching is that it can be used in conjunction with scheduling algorithms to reduce traffic congestion in the network.
  - Multi-path end-to-end load-balancing allows for several benefits, such as alleviating hotspots in the network, using underutilized network paths, and enhancing application level data transfer resiliency.
- **Examples of current activities**
  - OSG in-network caching pilot - This pilot involves deploying caches at UCSD, Caltech, and ESnet to support CMS workflows, and to contribute to the WLCG data lake for HL-LHC activities.
  - ESnet FPGA based network steering exploration.
- **Considerations**
  - Deploying the “3rd party” caching stacks requires a significant amount of coordination (e.g., getting the correct certs, balancing security concerns for superuser access, negotiating support models, etc) for on-going operations.
  - The deployment of multi-path capabilities may be specific to the network layer at which is it implemented. This solution may requiring “book-ending” the source and destination connection with the appropriate equipment for segmentation and reassembly.

Findings and Recommendations

- Physics needs to have the people to effectively integrate networking into larger Physics processes - whatever the technology stack of the day might be
- Standardization of orchestration APIs is needed for scaling automation/orchestration services in the larger ecosystem, vs. bespoke APIs for each resource.
- A unified statistics/telemetry platform across Physics sites would go far to facilitate end-to-end multi-domain traffic analysis.
- Data movement performance is an end-to-end capability. This requires all participating domains to offer congruent services and capabilities - in production as well as in development.
- Especially important are the transitions from research into production, which will require significant effort and should not be underestimated.

Fin