SC20 Advanced Networking For Data Intensive Science Experiments and Demonstrations

Joe Mambretti, Director, (j-mambretti@northwestern.edu)
International Center for Advanced Internet Research (www.icair.org)
Northwestern University
Director, Metropolitan Research and Education Network (www.mren.org)
Co-Director, StarLight (www.startap.net/starlight), Director, StarLight
International/National Communications Exchange Facility
(www.startap.net/starlight),
PI IRNC: RXP: StarLight SDX, Co-PI Chameleon, PI-iGENI, PI-OMNINet

JET Meeting
November 17, 2020
International Conference On High Performance Computing Networking, Storage and Analytics
November 9-20, 2020
Selected Applications

Compilation by Maxine Brown and Joe Mambretti
GRP: Services, Architecture, Technology

- Architecture: “Global Science DMZ”
- Services Optimized For Science Workflows
- High Performance Transport Over WANs For Large Capacity Data Streams E2E
- Enhanced Virtualization
- Highly Programmable
- Specialized Components (Software Stacks, Next Gen DTNs, Advanced APIs, Customized Devices, Federation, INT, P4, etc.)
- Interdomain Dynamic Provisioning
- Production Resources Adjacent to Testbeds
StarLight Global Research Platform Network
100 Gbps

Global Research Platform
Global Research Platform: Global Lambda Integrated Facility
Available Advanced Network Resources

Visualization courtesy of Bob Patterson, NCSA; data compilation by Maxine Brown, UIC.

www.glif.is
AutoGOLE Fabric 2020
StarLight – “By Researchers For Researchers”

StarLight: Experimental Optical Infrastructure/Proving Ground
For Next Gen Network Services
Optimized for High Performance
Data Intensive Science
Multiple 100 Gbps
(60+ Paths)
StarWave
100 G Exchange
World’s Most
Advanced Exchange
Multiple First of a Kind
Services and Capabilities

Abbott Hall, Northwestern University’s Chicago Campus
View from StarLight
StarLight SDX Overview

- This IRNC StarLight SDX Initiative Is Designing, Implementing, and Operating New Services For Global Data Intensive Sciences, Based On Emerging Next Generation Architecture and Technologies, Including Virtualization, Orchestration, Segmentation (Slicing), Software Defined Resources, Programmability and Customization.

- These Macro Trends Enable Exchanges To Be Agile Platforms For Dynamic Services Provisioning, Real-Time Responsiveness, and Distributed Control Over Core Resources, Including By Applications, Edge Processes and Devices.

This Project Is Transitioning Network Exchanges To Open Innovation Platforms
CENI NRE Map for SC20

List of NREs:
- NRE NRL B
- SL-3
- SL-4
- SL-5
- SL-7
- SL-9
- NRE-12
- SL-8
SC20 Experiments/Demonstrations Testbed
Design and implementation

Source: Se Young Yu, iCAIR
Jupyter Client for NSI OpenNSA Integration

Securely allow users to run NSI OpenNSA services (i.e. dynamically stitch layer 2/3 circuits based on technologies e.g. VLAN.)

- Features: Authentication, Authorization, Accounting
- Authentication: SSL/TLS operation authenticate both server and client. (certificates need to be exchanged prior unless using public certificates)
- Authorization: server will allow access users to request services on certain ports/VLANs based on user identification. i.e. Request command arguments will be parsed and authorized if allowed.
- Accounting: for future requirement.
- Additional Feature: Asynchronous (non-blocking operation allows multiple users to request services simultaneously)

Step 1: establish TCP connection with SSL/TLS client-server authentication
- Verify client certificate
- Verify server certificate

Step 2: client request service and server will parse the command arguments
- Authorize request and send back output/error
- Note: certain commands will return timely responses

Step 3: Finish and close the connection
- Close confirm

Source: Se Young Yu, iCAIR
SCinet DTN-as-a-Service @ SC20

For SC20, DaaS supports XNET Experiments, 12+ NREs: ROCE over WAN, P4 Experiments/Demonstrations, SENSE/openNSA integration, Kubernetes Federation, PCI-e Gen4 DTNs and Many More, Including The “Bring-Your-Own-Testbed Concept”
NRE09-GRP Service: DTNaaS for Petascale Sciences Data Transfer

As 10/09 Supports

NRE03, NRE04
NRE05, NRE06
NRE10, NRE11
NRE12, NRE13
NRE14

indis104s1

100G Science DTN testbeds Since 2011
NRE08-GRP Service: International P4 Experimental Networks (iP4EN)
Goal: Secure multi-domain resource sharing cross regional, national and international research platforms.

Solution: Admiralty. The software enables users to schedule workloads in a different cluster by federating the source and target clusters.

Participants:
- Global Research Platform (GRP)
- Pacific Research Platform (PRP)
- MREN Research platform (MRP)
- Towards National Research Platform (TNRP)
- PacificWave
- KISTI
BigData Express: Toward Predictable, Schedulable, High-performance Data Transfer

What Is BigData Express?
- A schedulable, predictable, and high-performance data transfer service
  - Peer-to-peer, scalable, and extensible data transfer model
  - Visually appealing, easy-to-use web portal
  - High-performance data transfer engine
  - On-demand provisioning of network paths with guaranteed QoS
  - Robust and flexible error handling
  - CILogon-based security
  - A rich set of REST APIs to support scientific workflows
- Project website: [http://bigdataexpress.fnal.gov](http://bigdataexpress.fnal.gov)

End-to-End Transfer Loop
- Application-aware network service
  - On-demand programming
  - Fast-provisioning of end-to-end network paths with guaranteed QoS
  - Distributed resource negotiation & brokering

Distributed, Peer-to-Peer Model
- Each site independently provides data transfer service
- Flexible data transfer federations

A High-performance Data Transfer Engine – mdtmFTP
- A pipelined I/O centric design to streamline data transfer
- Multicore-aware data transfer middleware optimizes use of underlying hardware
- Extremely efficient in transferring large amounts of small files

Scalable & Extensible Design
- Scheduler manages site resources through agents
- Use MQTT as message bus
- Extensible Plugin framework to support various data transfer protocols
  - mdtmFTP, GridFTP, XrootD

Miscellaneous
- Apache 2.0 License
- Docker release is available at [publicregistry.fnal.gov](http://publicregistry.fnal.gov)
- Manuals:
- Deployment and collaborations
- [KISTI](http://www.kisti.re.kr), [KRENET](http://www.kre.net), [MAX](http://www.max.org), [Ciena](http://www.ciena.com), [NCI](http://www.nic.gov), [STARBright](http://www.starbright.org), [ROCAF](http://www.rocaf.org), [REACH](http://www.reach.org), [KOSMOS](http://www.kosmos.org), [DA2CMO](http://www.da2cmo.org), [ORNL](http://wwwornl.gov), [BNL](http://www.bnl.gov), [LBNL](http://www.lbl.gov), [SLAC](http://www.slac.stanford.edu), [NRC](http://www.nrc.ca), [NIST](http://www.nist.gov), [NSF](http://www.nsf.gov), [DOE](http://www.energy.gov), [STARBright](http://www.starbright.org), [REACH](http://www.reach.org), [KOSMOS](http://www.kosmos.org), [ROCAF](http://www.rocaf.org), [DA2CMO](http://www.da2cmo.org), [NCI](http://www.nic.gov), [KISTI](http://www.kisti.re.kr), [KRENET](http://www.kre.net), [MAX](http://www.max.org), [Ciena](http://www.ciena.com)

[STARLIGHT](http://www.starlight.com)
ROBIN - PI Wenji Wu Fermi National Accelerator Laboratory

ROBIN (Rucio/BigData Express/SENSE)
A Next Generation High-performance Data Service Platform

Scientific Applications (e.g., LSST, CMS)

Rucio
Data Management Service

- Smart Namespace
- Easy Integration
- Peer-to-peer, Scalable, Extensible
- Software Defined-Networking (SDN) for End-to-End Virtual Guaranteed Network

BigData Express
High-performance Data Transfer Service

- Storage Support
- AA Support
- High Performance Data Transfer Engine
- CILogon-based Security

SENSE
Smart Network Service

- Consistency
- Proven Track Record
- End-to-end Performance Optimization
- A Highly Intuitive “Intent”-based Interface
ROBIN (Rucio/BigData Express/SENSE)
A Next Generation High-performance Data Service Platform
ROBIN Cross-Atlantic Testbed

StarLight site:
- DTN: `dtn110.sl.startap.net`, with several Intel NVMe drives for data storage, a 100GE Mellanox NIC for data transfer, and a 1G NIC for control.
- Head node: `165.x.x.157`, with a 1G NIC for control.

CERN site:
- DTN: `dtn01.cern.ch`, with a rotational disk for data storage, a 10GE Mellanox NIC for data transfer, and 1G NIC for control.
- Head node: `cixp-surfnet.cern.ch`, with a 1G NIC for control.
IRNC-OSG Global Fabric
PRP/TNRPs United States (Outside California) Nautilus Hypercluster Connects CENIC and 4 Regionals and 5 Internet2 Sites
Composable Platform as a Service
Instrument for Deep Learning & Visualization (COMPaaS DLV)

Using the Liqid Composable Infrastructure System for...

...on-demand post-processing, AI, deep-leaning, CFD, and visualization.

...interoperable Science DMZs and Data Transfer Nodes (DTNs) at a national/ international scale, using best strategies from the Pacific Research Platform (PRP) over the GLIF infrastructure.

...SAGE2™, the de facto operating system for managing Big Data on tiled-display walls, enabling local and remote collaborators to access, share and interact with a variety of digital media.

NSF award CNS-1828265 to UI/EEVLab for COMPaaS DLV
https://www.evl.uiuc.edu/compaas-dlv

SAGE2 is trademark of the University of Illinois Board of Trustees

Liqid: https://www.liqid.com/
PRP: https://prp.ucsd.edu
GLIF: https://www.glif.is
SAGE2: http://sage2.sagecommons.org/
CHAMELEON: A LARGE SCALE, RECONFIGURABLE EXPERIMENTAL INSTRUMENT FOR COMPUTER SCIENCE

Kate Keahey

Joe Mambretti, Pierre Riteau, Paul Ruth, Dan Stanzione
Next GRP Workshop

- Date: September 13-17, 2021
- Place: Innsbruck Austria, Co Located With eSCience Conference

- Global Research Platform Website:
- www.theglobalresearchplatform.net
THANKS!

Thanks to the NSF, DOE, NASA, NIH, DARPA
Universities, National Labs, International & Industrial Partners, and Other Supporters
"Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Networking and Information Technology Research and Development Program."

The Networking and Information Technology Research and Development (NITRD) Program

Mailing Address: NCO/NITRD, 2415 Eisenhower Avenue, Alexandria, VA 22314

Physical Address: 490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024, USA Tel: 202-459-9674, Fax: 202-459-9673, Email: nco@nitrd.gov, Website: https://www.nitrd.gov