Joint Engineering Team (JET) Meeting Minutes

National Coordination Office for Networking and Information Technology R&D (NCO/NITRD)
490 L'Enfant Plaza SW, Suite 8001, Washington, DC 20024
March 15, 2022, 12:00-2:00 p.m. ET
This meeting was held virtually

Participants

Nick Buraglio, ESnet	Paul Love, NCO/NITRD
Rich Carlson, DOE/SC	Joe Mambretti, StarLight/MREN
Bobby Cates, NASA/Ames	Linden Mercer, NRL
Basil Decina, NRL	Edward Moynihan, Indiana University
Ann Keane, NOAA	Aruna Muppalla, NASA/GSFC
Bill Fink, NASA/GSFC	Mark Mutz, NOAA
Jonah Keough, PNWGP/Pacific Wave	Glenn Ricart, US Ignite
Padma Krishnaswamy, FCC	Dan Taylor, Internet2
Michael Lambert, PSC/3ROX	Christian Todorov, Internet2
Tom Leman, Virnao	Chris Wilkinson, Intenet2

Proceeding: This meeting was chaired by Rich Carlson (DOE/SC).

I. **Action Items**:
- Internet2 and ESnet updates on their respective new networks.

II. **Review of the Minutes of the February 2022 meeting:** Two corrections were made.

III. **FABRIC and FAB Update – Tom Lehman**
   A. FABRIC and Fabric Across Boarders (FAB) are two separate but closely related NSF awards. The goal is to enable a completely new paradigm for distributed applications and internet protocols and services which has:
      a. A nation-wide fully programmable network – "everywhere programmable."
      b. Significant compute and storage at each node which are imbedded in the network, not just at the edges.
      c. Provides GPUs, FPGAs and network processors.
      d. Supports QoS over both dedicated optical 100G links and dedicated layer 2 capacity.
      e. Interconnects with national facilities: HPC centers, cloud & wireless testbeds, commercial clouds, the Internet, and edge nodes at universities and labs.
      f. Permits applications, protocols and services that can be run at any node in the network, not just at the network's edge or in the cloud
      g. Will be highly instrumented – the network as a scientific instrument.

B. Physical overview:
   a. 29 FABRIC Nodes (schedules/timelines have shifted due to supply chain issues)
      i. Development Phase: 3 Nodes
      ii. Phase 1: 16 Nodes
      iii. Phase 2: 10 Nodes + Terabit SuperCore
   b. 9 nodes co-located at ESnet6 Points of Presence
      i. Connected via dedicated 100G Layer 1 waves over DWDM across the new ESnet6 open line optical system. ESnet6's open line system is a key part of FABRIC
      ii. Some connections will be upgraded to the Terabit SuperCore during Phase 2
   c. 20 other nodes distributed across the R&E community at various regional networks, major cyberinfrastructure facilities, and university hosting sites. FABRIC is working to get as many connected over 100G Layer 1 as possible.
C. The FABRIC is a four year construction award and currently is about halfway through year three. Even though it is a construction award there are users on FABRIC even as it's being rolled out. As the number of deployed nodes grows this year FABRIC expects a pretty significant growth in the number of researchers making use of the system.
D. FABRIC has taken the approach of having two topologies – one for development and the other for production which alloys for the overlapping of deployment with development.
   a. Development consists of three sites: RENCI, the University of Kentucky and LLBL. It's rich enough environment to support the hardware and software development and connects most of those doing that work. With the design of FABRIC it's very easy to change topologies and migrate these over to the production network when appropriate.
   b. Production is growing as more nodes are deployed. Some is on layer 1 and some on layer 2 (over ESnet and Internet2). There have been delays in rollout as some optical modules have needed to be RMA'd. As these are renewed there will be a ring structure plus a StarLight<>Dallas cut-through direct path. As it evolves it will become a series of interconnect rings with spokes to many of the participants. Nodes and participants will each have a FABRIC rack. Current experimenters work with the sites that are available and can grow their experiments as more resources come online. There have been supply chain induced delays – the hardware for the nodes FABRIC expects to deploy in April/May have been on order since July 2021.
   c. In the final deployment push FABRIC will add several more nodes many of which will be at ESnet6's collocation facilities. That'll put them on the ESnet open line system so they will have DWDM waves. Several more edge sites as well. All the needed pieces for this last set should be ordered before the end of April. Ongoing supply chain issues will determine when these can be deployed – originally this was to be by mid-2023, but TBD.
E. FABRIC tries to work closely with the local RONs to engineer layer 1 paths wherever possible. A site may initially come on via a layer 2 connection and then be able to

migrate to desired layer 1 connection. Long term there are three connection options:

    a. Layer 1 DWDM over a RON to another FABRIC node

    b. Layer 2 Ethernet over a regional IP/layer 2 network to another FABRIC node

    c. Layer 2 Ethernet over a regional IP/layer 2 network to Internet2's AL2S or ESnet's OSCARS for onward transport to another FABRIC node.

    d. Regardless of the connection method the services FABRIC offers to the experiment will be the same. The only difference will be quality which is dependent on the type of path: layer 1 will have a great QoS while a shared layer 2 path may have issues outside FABRIC's control.

F. FABRIC has connections to ESnet, Internet2 (I2), various cloud providers via I2's cloud services, and other testbeds. It has IPv6 address space and an ASN so there will peer with ESnet and I2 using IPv6.

G. A FABRIC node is designed to interpose compute and storage into the path of fast packet flows. For example:

    a. Rack of high-performance servers (Dell 7525) with:

        i. 2x32-core AMD 7532 with 512G RAM

        ii. GPUs (NVIDIA RTX 6000, T4, A30), FPGA network/compute accelerators

        iii. Storage - the experimenter can provision 1TB NVMe drives in servers and a pool of ~250TB rotating storage at each site.

        iv. Network ports connect to a 100G+ switch, programmable through control software

    b. Reconfigurable network interface cards:

        i. FPGAs (U280 XILINX with P4 support)

        ii. Mellanox ConnectX-5 and ConnectX-6 with hardware off-load

        iii. Multiple interface speeds (25G, 100G, 200G+(future)

    c. Kernel Bypass/Hardware Offload

        i. VM/Containers sized to support full-rate DPDK for access to programmable NIC, FPGA, and GPU resources via PCI pass-through.

    d. Besides using the various PCI cards, experiments can bring their own hardware to support their local experiment.

    e. An experimenter can request a bare metal server as well as the more common PCI pass-through for VMs and containers

H. Measurements

    a. High precision measurements are a key to FABRIC being a scientific instrument

    b. Hardware in a FABRIC rack to support this include:

        i. GPS-disciplined clock source with PTP at most sites. (Since FABRIC is capable of very high precision measurement this is very important. Most sites have been able to solve placing an antenna (network carrier spaces are problematical)).

        ii. NICs capable of accurate packet sampling/timestamping

        iii. Programmable port mirroring

        iv. PDUs able to measure power

        v. Optical layer measurements when there are layer 1 connections

        vi. CPU, memory, and port/interface utilization measurements

    c. Measurements can be made inside an experiment and also outside in the node's infrastructure.

I. BYOE hardware
   a. Three possible ways to do BYOE
      i. A PCI card that is integrated into a FABRIC node
      ii. Standalone server
      iii. Standalone switch.
   b. BYOE is welcomed with the usual constraints; power, space, management & data plane port availability and PCI slots if needed.
   c. BYOE can be integrated into the FABRIC control and measurement frameworks.

J. An experiment is contained in a slice
   a. Slices consist of slivers of individually programmable or configurable resources
   b. Slices can change over time – they can grow or shrink, adding or shedding resources under programmatic control
   c. Slice topologies can be connected across FABRIC via:
      i. Custom layer 2 using the underlying MPLS-SR
      ii. Over the routable IPv6/IPv4 layer within FABRIC
      iii. External layer 3 routed access via IPv6 public space
   d. Basic sliver classes
      i. Nodes - can include a selection of PCI-pass through devices
      ii. Links – layer 2 or layer 3 with or without QoS
      iii. Measurement points - inside and outside the slice
   e. An experiment (slice) can run in isolation within a FABRIC node or with intra-FABRIC connectivity via routed layer 3 or dedicated layer 2 paths. A slice can also make use of FABRIC's external connections and peerings - layer 3 peering, layer 2 services peering or public cloud connections via I2's cloud services.

K. Portal
   a. A beta version of the FABRIC portal is up
      i. Provides system services (credential management, project management, and user information management)
      ii. Jupyter Hub
      iii. Control framework
      iv. Measurement framework
      v. Will support both basic researchers and those who are advanced in needs
   b. Integrated with CILogon and COmanage.

L. FABRIC Across Boarders (FAB)
   a. To enable global experimentations
   b. Four international FABRIC nodes; University of Tokyo, University of Bristol, University of Amsterdam, and CERN
   c. FABRIC works with a variety of facility and network partners.
      i. In Europe FABRIC will use the NEA^{3}R links and perhaps others – discussions are underway.
      ii. In Asia-Pacific the path will probably be over TransPAC and SINET. KERONET may also help with connectivity

iii. To reach South American experimenters from the node at FIU AmLight,
                ANSP and RedCLARA will carry the traffic
     M. Questions?
          a. Will the European sites use GPS or Galileo?
              So far, the discussions have been about using GPS. Unclear what the implications
              would be in using Galileo.
          b. For those sites, such as network carrier buildings, where getting GPS is an issue,
              what's the fall back?
               i. The plan is to use a GPS PTP feed from the local RONs which have layer 1
                   waves to the FABRIC node in the carrier facility. The feed will be in a high
                   priority queue, with high QoS. FABRIC is pioneering this approach. There
                   should be little to no jitter. Any latency can be corrected for. The
                   approach remains to be tested.
          c. FAB node sequencing: The first will be CERN probably followed by Bristol with
              Amsterdam and Tokyo the final two. All are a bit uncertain due to supply chain
              and customs issues.
     N. More information:
          a. https://whatisfabric.net and info@fabric-testbed.net
          b. FABRIC Software: https://github.com/fabric-testbed
          c. Beta Tester request form:
              https://whatisfabric.net/get-involved/beta-testers-request

IV. **Discussion of the JET's tasking on tools to help with inter-domain issues – Joe Breen (via email)**
This is a community project to collect shared data from all who are willing to share.
The related, live map is at: https://www.globalresearchmap.org/
   • Progress continues albeit at a slow pace due to higher priority projects for those
      involved.

V. **Operational network security roundtable**
     A. ESnet (Nick Buraglio): One of ESnet6's deliverables is a filtering system, SCRAM, which
         ESnet's been working on for a couple of years. The base framework has built in an entry
         level set of features. It allows ESnet to do network wide blocking that does not require
         touching every device within the network.
         It permits ESnet to do very efficient black hole filtering across the network. Eventually it
         will include such things as enabling extensive flow spec filtering. The pieces are in place
         to permit that to be easily added to SCRAM.
         Within two hours of the entry level service being put into production ESnet had a use
         case to leverage it and it worked perfectly.
     B. 3ROX (Michael Lambert): Team Cymru has release v 2.0 of its Unwanted Traffic Removal
         Service (UTRS). The new version has added support for IPv6. It also has some
         improvements by supporting fewer specific matches for black holing traffic.

VI. **Network roundtable**
   A. ESnet (Nick Buraglio):
      a. The ESnet6 brownfield upgrades are continuing. Significant progress has been made with the ESnet6 deliverables. ESnet expects to have all completed by the end of the month.
      b. A significant amount of work is being done at the Department of Energy level on the IPv6-only mandate. For ESnet the IPv6-only management network is in place. ESnet is getting ready to run some services using IPv6-only as a trial in a data center.
      c. Over the last six months or so ESnet has developed an orchestration testing platform where the ESnet orchestration team could test their workflows. It'll be described at TNC22. This platform, ROVER, allows a developer to bring up a corner of ESnet6 with five Nokia routers, an IPAM instance, a NSO instance, an instance of ESnet's internal data base plus a couple of other items. A developer can then run the workflow through to validate or adjust and rerun. When done its all torn down. The talk at TNC will include a demo.
         n.b.: TNC has a no cost registration for remote access which should stream this (and all other) session.
   B. International Networks – Indiana University (Ed Moynihan):
      a. NEA^{3}R: Links are stable. A new MOU/partnership agreement for ANA is under development. That will be announced at TNC, if not before.
      b. Pacific: Both the TransPAC and Guam<>Singapore links are stable. The routing WG International Networks/Indiana University is helping to lead held a meeting during APAN last week.
   C. NASA/GSFC (Bill Fink): Continuing to work with Joe Mambretti (StarLight) and Linden Mercer (NRL) on plans for SC22.
   D. NOAA (Mark Mutz): N-Wave continues to work on its core router upgrades to Juniper's PTX boxes. By the mid-June, the backbone links will have also been upgraded to 400G using Internet2.
   E. NRL (Linden Mercer): With StarLight and GSFC, NRL is working on plans for SC22.
   F. Pacific Wave (Jonah Keogh):
      a. Pacific Wave (PW) has had its first customer upgrade to a 200G connection (NII in Los Angeles, CA).
      b. PW's route servers (RS) are ready for participants' connections. PW did a soft launch of this service but is ready for customers interested in more easily setting up BGP peerings across the exchange point. The RS will also permit the enforcement of MANRS rules that it can't over a layer 2 exchange.
      c. PW is continuing with its West Coast backbone upgrade. It's had some issues with Ciena gear needing to be RMA'd upon arrival. These have been resolved and the upgrade is going forward.
   G. PSC/3ROC/XSEDE (Michael Lambert): Following up on Tom's update on FABRIC, 3ROX is working to pull together all its requirements to get to FABRIC. The GPS antenna is at the top of 3ROX's list on how to make that work.

H. US Ignite (Glenn Ricart):
   a. The Smart Cities Conference will be held April 4-7 in Columbus, OH. A number of white papers will be released in conjunction with it on networking, local networking, and municipal networking. There's increased interest currently as there is $42B available to support connecting the unconnected.
   b. US Ignite (USI) did an NSF funded pilot project with seven communities across the US. Six of the seven are now in operation. (The seventh, on doing a neighborhood fiber project, ran into several issues – permits, and other non-technical slowdowns.) Of the first two up one was a demonstration of how to quickly recover when a forest fire burned up the fiber connection for the small community of Blue River, OR – put up a tower for over the air connectivity. The other, with the help of the Missouri University of Science and Technology, built a tower in Clinton County, MO, to give line-of-sight to the homes in a small town in the county. One discovery was that, due to the distances involved, outdoor antennas were needed.
   c. In each community USI has worked closely with local community organization and anchor institutions to make sure that each community has the resources, training, etc. to use the new connections. This has really paid off.

VII. **Exchange Points Round Table**
   A. StarLight (Joe Mambretti):
      a. StarLight (SL) is working with NRL and GSFC to develop a Joint Big Data Testbed (JBDT) facility in McLean, VA. A 100G circuit has just come up between JBDT there and SL. SL is working with Tom Lehman to get a 100G FABRIC connection into the JBDT facility as well. SL is also working on a possible 400G connection over ESnet with Tom Lehman's help. All this is in part preparation for SC22 in November.
      b. Related to SC22: April 29 due date for any WAN requirements for a Network Research Exhibition. Full details can be found at:
         https://sc22.supercomputing.org/scinet/network-research-exhibition/
      c. SL participated in the SupercomputingAsia (SCA) conference. In the months leading up to SCA SL had helped develop a 100G international test bed for the conference's Data Mover challenge. SL also participated in the challenge and won the long distance high performance award and the Data Mover Challenge overall award.
      d. SL helped organized a Global Research Platform workshop during SCA. Several large science projects were discussed – the Square Kilometer Array, the Pierre Auger Observatory which has a large synoptic survey telescope, and the high luminosity experiment at CERN's LHC. Speakers from each were invited to talk about their developing architecture services technology infrastructure. GÉANT gave a presentation on its Next Generation Network plans. There was also a presentation of the 2x100G plans for the Netherlands<>KAUST <>Singapore connection.

B. Ames (Bobby Cates): Nothing new today for the exchange.
C. PNWGP (Jonah Keogh): No updates today.

Meetings of Interest 2022
Note: Meetings whose format has changed have been updated.
Mar 19-25   IETF 113, Vienna, Austria, hybrid
Apr 24-27   ARIN 49, Nashville, TN
Jun 6-8   NANOG 85, Montréal, QC, Canada, hybrid
Jun 13-17   TNC22, Trieste, Italy, primarily in-person with a basic remote option
Jul 10-14   PEARC22, Boston, MA
Jul 23-29   IETF 114, Philadelphia, PA
Aug 22-26   APAN54, China, primarily virtual with possibly limited local attendance
Sep 20-22   The Quilt Fall Meeting, South Minneapolis, MN
Oct 10-11   Global Research Platform Workshop, Salt Lake City, UT
Oct 17-19   NANOG 86, Hollywood, CA
Oct 20-21   ARIN 50, Hollywood, CA
Nov 13-18   SC22, Dallas, TX
Nov 5-11   IETF 115, London, UK
Dec 5-8   Internet2 Technology Exchange, Denver, CO

Next JET meetings
Note: It is anticipated that JET meetings will remain virtual for the foreseeable future
Apr 19, 2022   12-2 p.m. ET
May 17, 2022   12-2 p.m. ET
Jun 21, 2022   12-2 p.m. ET