Notes from LHCONE P2P experiment discussions Ann Arbor meeting, September 16, 2014

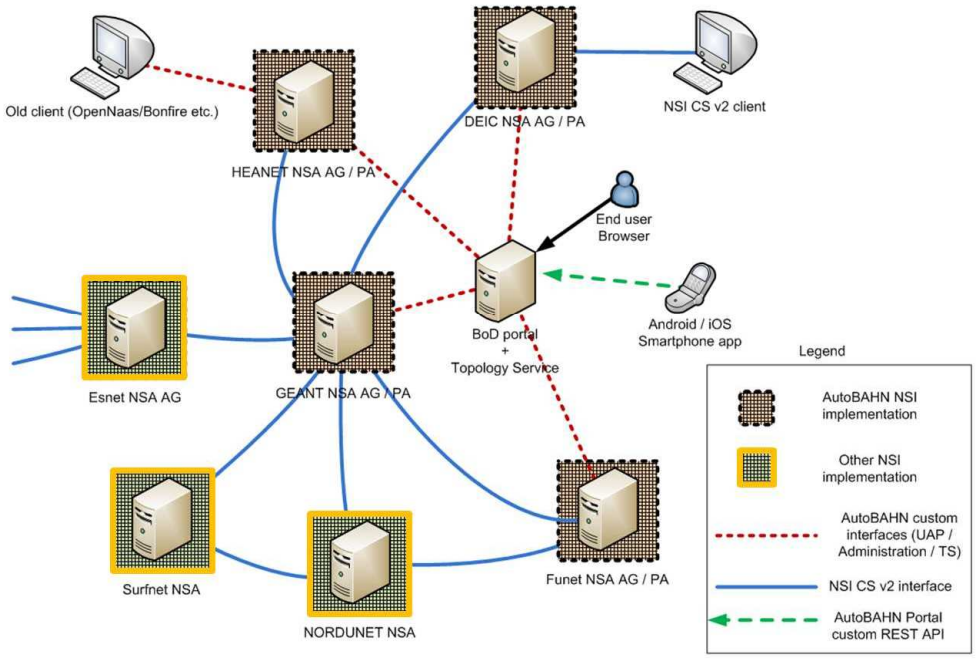
(https://indico.cern.ch/event/318811/)

W. E. Johnston

1 Update of GÉANT NSI implementation: AutoBAHN

Tangui Coulouarn (Dante)

- AutoBAHN 3.1 will be deployed on backbone Sept. 2014 and AutoBAHN 2.4 (IDCP) will be decommissioned
- Deployed in some NRENs:
  - HEAnet
  - FUnet
  - DeIC
  - GRnet (test)
  - JANET(test)
  - DFN (test)
  - Carnet (test)
  - GARR – no
  - RENATER – static link to GEANT PoP
- Each instance functions both as a Request and Provider Agent RA/PA and simple Aggregator (AG).
  - Simple AG means that it follows the CHAIN model: It is capable of sending the request to its own PA and the next PA on the reservation path



2 NORDUnet

Tangui Coulouarn (Dante)

- Well working implementation (OpenNSA), that is used on multiple sites
  - Close collaboration with SURFnet and GÉANT
  - Well tested and connected to SURFnet and GÉANT
- Available on four edge nodes, MPLS tunnel over production network (Amsterdam, London, CPH, Helsinki)
- Full integration into production infrastructure
  - NOC support / handover
  - Coordination with NSI group in AA & Topology
  - Committed to launching an NSI/BoD production service
- Used in production for real circuits
  - CSC (Finland) to SurfSARA (Netherlands) for ELIXIR (European infrastructure for biological information, supporting life science research and its translation to medicine, agriculture, bioindustries and society)
    - using AutoBAHN (Funet), OpenNSA (NORDUnet) and BoD (SURFnet)

3 ESnet – OSCARS

John MacAuley, ESnet

- OSCARS provides Ultimate Provider (uPA) functionality exposing ESnet network resources to the NSI community (nsi-bridge).
- Fully supports the NSI CS v2 provider specification (revision 117)
- The “nsi-safnari” aggregator is jointly developed by ESnet and SURFnet.
  - The “nsi-safnari” aggregator is jointly developed by ESnet and SURFnet.
  - It is a 100% pure aggregator NSA with no associated uPA capabilities
- Deployed in three sites with more on the way

4 Internet2 and MAN LAN

Dale Finkelson

- OSCARS is not in Internet2 yet
- MAN LAN and WIX are running OSCARS

5 StarLight

Joe Mambretti

- StarLight is enabled
- Much of Joe’s talk was an interesting discussion of AutoGOLE and SDX experiments

6 NetherLight

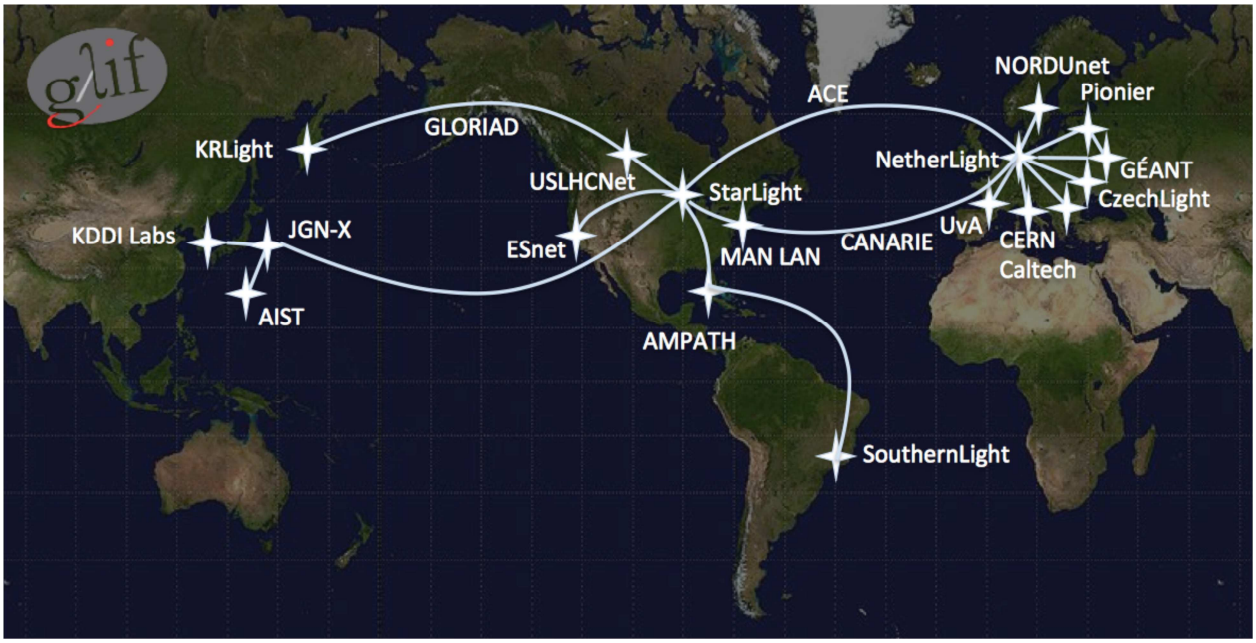
Gerben van Malenstein

- Running Automated GOLE in production since September 12, 2014 (fully compliant to NSI Connection Service v2.0).
- Running nsi-safnari aggregator to enable users to make reservations between any pair of endpoints in the world.
- AutoGOLE: direct data plane connections to NORDUnet, GÉANT, CESNET, SURFnet, MAN LAN, and StarLight.

7 Proposal for using AutoGOLE for LHCONE P2P experiments

Gerben van Malenstein (Surfnet), presented by Tangui Coulouarn (Dante)

- Proposal: Use AutoGOLE To demonstrate a working implementation/solution of the LHCONE Point2Point Experiment with a number of LHC sites
- The Automated GOLE was moved into produc9on on Friday September 12, 2014. We do expect further necessary improvements throughout the fabric, but need user involvement to evolve

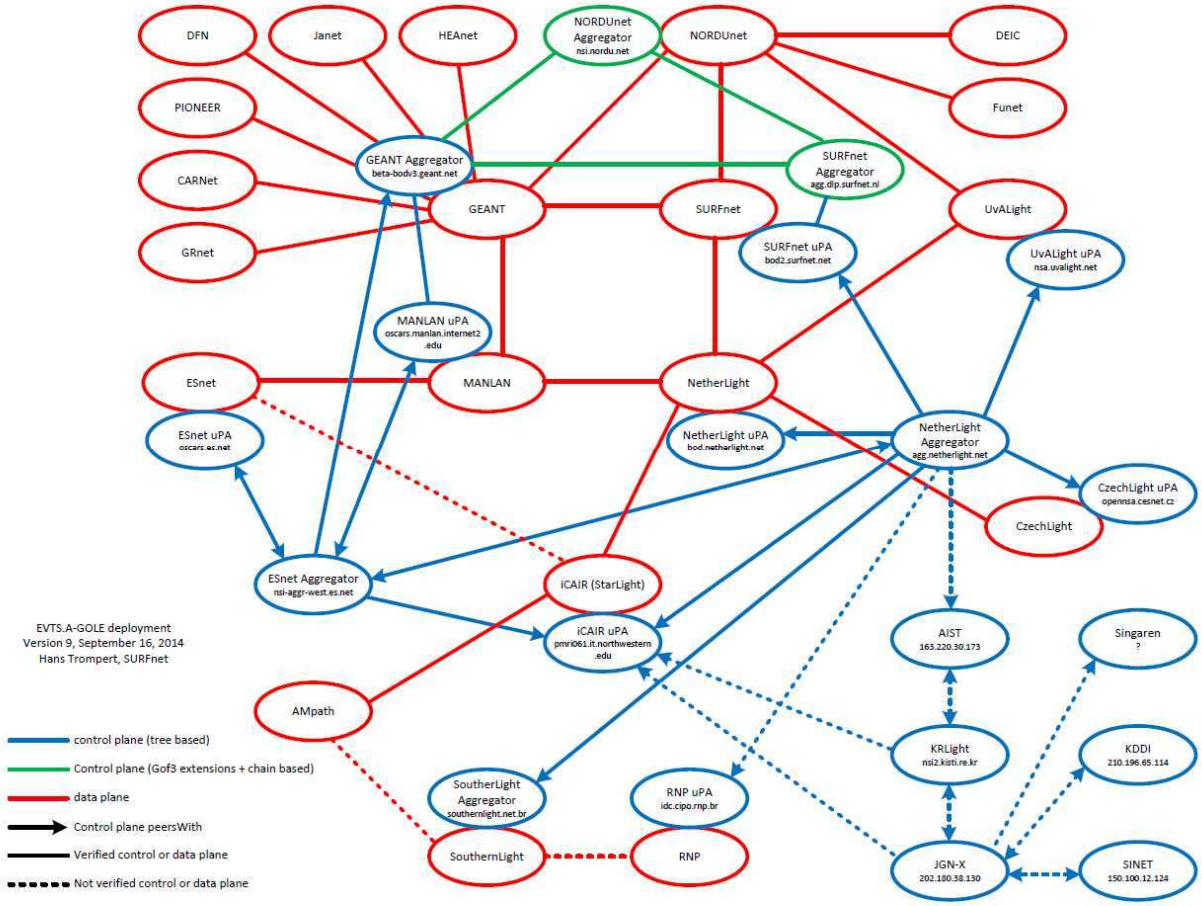


- For the LHCONE Point2Point Experiment, the Automated GOLE offers to provide a starting point to conduct real traffic experiments between a small number of LHC sites
- Two - three sites per continent to start with
  - For Europe these could be DE-KIT, NDGF and SURFsara
  - US sites may volunteer, as well as for e.g. Latin America and Asia.
- After the initial LHC sites are identified to join the AutoGOLE, a joint meeting will be setup to
  - gather input/expectations from the LHC sites
  - provide information on the AutoGOLE fabric and NSI
  - create a plan together
  - guide people through the process of connecting
- LHC sites would need to connect on the data plane to the nearest AutoGOLE participant

- Sites would use NSI-CS v2 to use the NSI connection service
- AutoGOLE only supports L2 – VLAN – connections
- Before SC'14
  - LHC sites to join P2P experiment on AutoGOLE are identified
- End of 2014
  - LHC sites are connected to nearest AutoGOLE participant on the data plane
- 2015Q1
  - First implementations and testing with middleware (NSI) by LHC sites
- 2015Q2
  - Experiment results and conclusion with the option to continue the effort by adding more LHC sites

wej-cmt: Gerben and I had a discussion of this proposal before the meeting and while I think that this is a basically good idea, I did raise a couple of points that troubled me:
If this experiment is a success and the P2P experiments lead to production-like service from the Experiments point of view and if this is done on R&D infrastructure, there could be considerable difficulties moving the service to production links supported by the network providers just because of physical resource constraints and because the network operations folks have not been involved.
Gerben makes the point that AutoGOLE has moved to production status, but what is the provenance of the underlying links? Are they committed to production service as well like the links of the network providers' production networks? Are the providers' operations staff involved?
None-the-less, I think that we should move forward with this because in AutoGOLE there is a community of support for this sort of thing that is not really in LHCONE.

Tangui comments: A sufficient number of networks have now deployed interoperable and interconnected NSI compliant services as represented by the AutoGOLE fabric diagram made by Hans. Testing, debugging, etc. are not over, we need static circuits to reach the services in many cases, etc. But the network providers are ready to start serving some early users. And the AutoGOLE can be used for that.



8 Discussion of the status of AuthN/AuthZ and its implications

Notes from Tangui Coulouarn

- The discussion on AAI brought several results. There seems to be a consensus on the fact that the network providers will not try to implement transit and SDP policies at least in a first phase. AAI should happen and be enforced mostly at the end sites. Copy-pasted from an email from John:
- "Two initial requirements for investigation:
  - Endpoint authorization based on project specific credentials for LHCONE based on the sub-projects (ATLAS, CMS, etc). This is done to provide access to their existing project infrastructure. Credentials are X.509 certificates with specific signed project attributes. Reservation requests will probably being coming in from the workload manager and not individual users, so a smaller subset of certificates will be used.
  - Project-based bandwidth restrictions. Want to make sure a single project is not consuming all the available resources. Explained this could not be enforced network wide based on the distributed control plane, but could be done on an end-site basis. Suggestion was made to introduce a centralized collection point to determine network wide usage. Interesting and perhaps a feature for a NSI monitoring service?"

This will be discussed in the NSI WG meeting in Sweden the week after the Ann Arbor meeting.

9 NSI Update

John MacAuley

- NSI Connection Service v2.0 has standard status
  - Group is actively building 2.0 errata based on developer feedback
- GLIF NSI Implementation Taskforce has been addressing deployment issues and providing feedback and new contributions into the NSI Working Group for standardization
- NSI Values (goals)
  - Complete decoupling of signaling plane from the data plane.
  - Deployment of Network Service Agents with no network associations.
  - Ability to perform centralized path finding with a complete view of the inter-domain topology.
  - Facilitate advanced network resource workflows for network aware applications.
  - Support for both tree and chain based signaling as a deployment option

10 NSI implementation testing

John MacAuley

- Base integration testing of NSI CSv2 protocol (r117) and client authenticated TLS is completed for implementations in North America, Europe, and Scandinavia.
- Other implementations are coming on-line daily
  - AMpath, SouthernLight, and RNP are being tested now.
  - Waiting for implementations in Asia to be made available for testing.
- What next?
  - Designers are busy fixing bugs in implementations, specifically around a clean and consistent connection lifecycle.
  - Consistent and useful error feedback for reservation failures.
  - Path finding feedback to guide reservation retries (using error feedback).
  - Individual implementation and control plane stress testing.

11 NSI Authentication and Authorization

John MacAuley

- Trust in the control plane
  - The control plane is built up through a series of NSA peering agreements (may or may not follow data plane peering).
  - Client authenticated TLS is used to authenticate peer NSA, as well as ensure the integrity and confidentiality of the messages traveling through the control plane.
  - Control plane security is based on transitive trust: I trust my neighbors and the neighbors they trust.
- User access to control plane
  - uRA must authenticate originating user using locally defined authentication schemes.
  - Identity information of the originating user can be added to the NSI message header and passed to peer NSA along with the reservation request.

- NSA along the reservation path can utilize identity information if needed for local policy enforcement.

- Authorization policies
  - uRA must authenticate originating user using locally defined authentication schemes.
  - Identity information of the originating user can be added to the NSI message header and passed to peer NSA along with the reservation request.
  - NSA along the reservation path can utilize identity information if needed for local policy enforcement.
- Policy impacts on path finding can be considerable and poorly understood by the end-user as they cannot see the policies of intermediate networks

12 CMS data transfer challenges

Azher Mughal

- CMS Software Components Primer
  - PhEDEx
    - Book keeping for CMS Data Sets. Knows the End points and manages high level aspects of the transfers (e.g. file router).
  - FTS
    - Negotiates the transfers among end sites/points and initiates transfers through the GridFTP servers.
  - SRM
    - Selects the appropriate GridFTP Server (mostly round-robin).
  - GridFTP
    - Actual workhorse or grid middleware for the transfers between end sites. Or, an interface between the storage element and the wide area network.
- More interesting information in the talk about the issues currently being faced in high-speed, long-distance data transfers

13 Candidate end sites for P2P experiment

Shawn McKee, ATLAS

- The AutoGOLE testbed involves all of the networks involved in the LHCONE P2P experiment
- What sites are connected to networks with P2P enabled?
- Caltech; U. Mich.; U. Texas, Austin; Vanderbilt
  - Need to identify people to conduct testing
- Need sites in Europe

14 LHCONE AUP – Belle-II wants to join LHCONE

Edoardo Martelli (CERN), Dr. Stefan Lueders (CERN)

wej-cmt: I include this discussion here because this applies to LHCONE P2P as well as to the VRF

This discussion follows from the request by the Belle-II at KEK experiment to join LHCONE. Initially there was discussion about Belle-II setting up its own VRF infrastructure, but analysis of

the resources showed that there was a high degree of commonality with LHC resources. In other words, many sites participate in both the LHC and Belle-II experiments, and the compute and data resources are shared between the two. This situation makes it difficult to partition the resurces among several VRFs, hence the request to join LHCONE.

- CERN expectations:
  - Resources connected to LHCONE must be limited to LHC/HEP resources
    - This is mostly accomplished by trusting the end sites to comply
  - "LHC/HEP resources" must not be general purpose systems
  - Must be able to do "rational" traffic analysis [within LHCONE]
  - "LHC/HEP resources" must have logging (e.g. syslog) enabled for security investigations
  - The EGI/WLHC security policy is described in
    - https://edms.cern.ch/file/428035/7/SecurityIncidentResponse-v3.2a.pdf
    - https://edms.cern.ch/file/428008/5/Security_Policy_V5.7a.pdf
    - https://edms.cern.ch/file/819783/2/GridSiteOperationsPolicy-v1.4a.pdf
- The Belle Computing Steering Group should coordinate with the WLCG Management Board
- Any Belle-II sites connecting to LHCONE should have comparable infrastructure, e.g. perfSONAR
- Belle-II had a formal computing MoU similar to the LHC MoUs. Malachi Schram (DOE Lab PNNL) had provided this to Edoardo to look at this the context of this duiscussion.
- Sho Suzuki points out that KeK is already a member of EGI

Edoardo is coordinating the modification of the LHCONE AUP document:

https://twiki.cern.ch/twiki/bin/view/LHCONE/LhcOneAup