ANSE-RELATED PROJECTS: LHCONE, DYNES AND OTHERS
AN OVERVIEW

Artur Barczyk/Caltech
2nd ANSE Collaboration Workshop
Snowmass on the Mississippi
Minneapolis, July 2013
LHCONE: INTRODUCTION
LHCONE Introduction

- In brief, LHCONE was born to address two main issues:
  - ensure that the services to the science community maintain their quality and reliability
  - protect existing R&E infrastructures against potential “threats” of very large data flows
- LHCONE is expected to
  - Provide some guarantees of performance
    - Large data flows across managed bandwidth that would provide better determinism than shared IP networks
    - Segregation from competing traffic flows
    - Manage capacity as # sites x Max flow/site x # Flows increases
  - Provide ways for better utilization of resources
    - Use all available resources
    - Provide Traffic Engineering and flow management capability
    - Leverage investments being made in advanced networking
Current activities split in several areas:

- **Multipoint connectivity through L3VPN**
  - Routed IP, virtualized service

- **Point-to-point dynamic circuits**
  - R&D, targeting demonstration this year

- Common to both is logical separation of LHC traffic from the General Purpose Network (GPN)
  - Avoids interference effects
  - Allows trusted connection and firewall bypass

- **More R&D in SDN/Openflow** for LHC traffic
  - for tasks which cannot be done with traditional methods
LHCONE: ROUTED IP SERVICE
Based on Virtual Routing and Forwarding (VRF)

BGP peerings between the VRF domains

Currently serving 44 LHC computing sites
Current logical connectivity diagram:
Inter-domain connectivity

- Many of the inter-domain peerings are established at Open Lightpath Exchanges
- Any R&E Network or End-site can peer with the LHCONE domains at any of the Exchange Points (or directly)

<table>
<thead>
<tr>
<th></th>
<th>MANLAN</th>
<th>StarLight</th>
<th>WIX</th>
<th>NetherLight</th>
<th>CERNLight</th>
</tr>
</thead>
<tbody>
<tr>
<td>GEANT</td>
<td>⭐️</td>
<td>⭐️</td>
<td>⭐️</td>
<td>⭐️</td>
<td>⭐️</td>
</tr>
<tr>
<td>NORDUnet</td>
<td>⭐️</td>
<td></td>
<td></td>
<td></td>
<td>⭐️</td>
</tr>
<tr>
<td>Internet2</td>
<td>⭐️</td>
<td>⭐️</td>
<td>⭐️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ESnet</td>
<td>⭐️</td>
<td>⭐️</td>
<td>⭐️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CANARIE</td>
<td>⭐️</td>
<td>⭐️</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASGC</td>
<td>⭐️</td>
<td></td>
<td></td>
<td>⭐️</td>
<td></td>
</tr>
</tbody>
</table>
LHCONE: POINT-TO-POINT SERVICE

PATH TO A DEMONSTRATION SYSTEM
Dynamic Point-to-Point Service

• Provide reserved bandwidth between a pair of end-points
• Several provisioning systems developed by R&E community: OSCARS (ESnet), OpenDRAC (SURFnet), G-Lambda-A (AIST), G-Lambda-K (KDDI), AutoBAHN (GEANT)
• Inter-domain: need accepted standards
• OGF NSI: The standards Network Services Interface
• Connection Service (NSI CS):
  – v1 ‘done’ and demonstrated e.g. at GLIF and SC’12
  – Currently standardizing v2
GLIF and Dynamic Point-to-Point Circuits

- GLIF is performing regular demonstrations and plugfests of NSI-based systems
- Automated-GOLE Working Group actively developing the notion of exchange points automated through NSI
  - GOLE = GLIF Open Lightpath Exchange

This is a R&D and demonstration infrastructure!

Some elements could potentially be used for a demonstration in LHCONE context
Intended to support bulk data transfers at high rate
Separation from GPN-style infrastructure to avoid interferences between flows

LHCONE has conducted 2 workshops:
- 1st LHCONE P2P workshop was held in December 2012
  - https://indico.cern.ch/conferenceDisplay.py?confId=215393
- 2nd workshop held May 2013 in Geneva
  - https://indico.cern.ch/conferenceDisplay.py?confId=241490

(Some) Challenges we face:
- multi-domain system
- edge connectivity – to and within end-sites
- how to use the system from LHC experiments’ perspective
  - e.g. ANSE project in the US
- manage expectations
Point-to-point Demo/Testbed

- Demo proposed at the 2nd workshop by Inder Monga (ESnet)
  1) Choose a few interested sites
  2) Build static mesh of P2P circuits with small but permanent bandwidth
  3) Use NSI 2.0 mechanisms to
     - Dynamically increase and reduce bandwidth
     - Based on Job placement or transfer queue
     - Based or dynamic allocation of resources
- Define adequate metrics!
  - for meaningful comparison with GPN or/and VRF

- Include both CMS and ATLAS
- **ANSE is a key part in this – bridging the infrastructure and software stacks in CMS and ATLAS**

- Time scale: TDB (“this year”)
- Participation: TDB (“any site/domain interested”)
LHCONE: SDN/OPENFLOW

OTHER R&D ACTIVITIES
• **Software Defined Networking (SDN):** Simply put, *physical separation of control and data planes*

• **Openflow:** a protocol between controller entity and the network devices

• The potential is clear: a network operator (or even user) can write applications which **determine** how the network behaves

• E.g. centralized control enables efficient and powerful optimization (“traffic engineering”) in complex environments
Discussed the potential use case: SDN/Openflow could enable solutions to problems where no commercial solution exists

Identify possible issues/problems Openflow could solve, for which no other solution currently exists?

- Multitude of transatlantic circuits makes flow management difficult
  - Impacts the LHCONE VRF, but also the GPN
  - No satisfactory commercial solution has been found at layers 1-3
  - Problem can be easily addressed at Layer2 using Openflow
  - Caltech has a DOE funded project running, developing multipath switching capability (OLiMPS)
  - We’ll examine this for use in LHCONE

- ATLAS use case: flexible cloud interconnect
  - OpenStack deployed at several sites.
  - Openflow is the natural virtualisation technology in the network. Could be used to bridge the data centers
• Initiated by Caltech and SARA; now continued by Caltech with SURFnet
  – Caltech: OLiMPS project (DOE OASCR)
    • Implement multipath control functionality using Openflow
  – SARA: investigations of use of MPTCP
• Basic idea: Flow-based load balancing over multiple paths
  – Initially: use static topology, and/or bandwidth allocation (e.g. NSI)
  – Later: comprehensive real-time information from the network
    (utilization, topology changes) as well as interface to applications
  – MPTCP on end-hosts
• Demonstrated at GLIF 2012, SC’12, TNC 2012
• Started with local experimental setup
  – 5 link-disjoint paths, 5 Openflow switches
  – 1 to 10 parallel transfers
  – Single transfer (with multiple files) takes approximately 90 minutes
  – File sizes between 1 and 40 GByte (Zipf); 500 GByte in total
  – Exponentially distributed inter-transfer waiting times

• Compared 5 different flow mapping algorithms

• Best performance: **Application-aware** or number-of-flows path mapping
OLiMPS and OSCARS dynamic circuits

OLiMPS/OSCARS Interface

- User (or application) requests network setup from OLiMPS controller
- OLiMPS requests setup of multiple paths from OSCARS-IDC
- OLiMPS connects OpenFlow switches to OSCARS termination points, i.e. VLANs
- OLiMPS transparently maps the site traffic to the VLANs
DYNES

DYnamic NEtwork Services
DYNES is an NSF funded project to deploy a cyberinstrument linking up to 50 US campuses through Internet2 dynamic circuit backbone and regional networks
  – based on ION service, using OSCARS technology

PI organizations: Internet2, Caltech, UoMichigan, Vanderbilt

DYNES instrument can be viewed as a production-grade ‘starter-kit’
  – comes with a disk server, inter-domain controller (server) and FDT installation
  – FDT code includes OSCARS IDC API reserves bandwidth, and moves data through the created circuit
    • “Bandwidth on Demand”, i.e. get it now or never
    • routed GPN as fallback

The DYNES system is naturally capable of advance reservation

ANSE: We need the right agent code inside CMS/ATLAS to call the API whenever transfers involve two DYNES sites

July 31, 2013
DYNES High-level topology

Internet2 ION

U of Colorado-Boulder
U of Nebraska-Lincoln
U of Oklahoma
U of Michigan
U of Wisconsin-Madison
U of Chicago
UIUC
Indiana University
U of Iowa
U of Southern Illinois

FRGP
GPN
MREN/CIC OmniPoP
ESnet SDN
NOX

Boston University
Harvard Tufts

USLHCnet

GEANT

CERN

U of Pennsylvania
Rutgers University
U of Delaware

AutoBAHN

U of California-Santa Cruz
U California San Diego

CENIC
LEARN
FLR AMPATH

FIU
U of Florida

SOX
MCNC
MAGPI
MAX

RNP

State University of Sao Paulo
State University of Rio de Janeiro: Alberto Santoro
Cerro Tololo Inter-American Observatory (CTIO)
Atacama Large Millimeter Array (ALMA)
Pierre Auger Observatory
Academic Network of Sao Paulo (ANSP)
Rede Nacional de Ensino e Pesquisa (RNP)
Red Universitaria Nacional (REUNA) de Chile
DYNES Sites

DYNES is ramping up to full scale, and working toward routine Operations in 2013

DYNES is extending circuit capabilities to ~40-50 US campuses

Intended as integral part of the point-to-point service in LHCONE
FDT integrates OSCARS IDC API to reserve network capacity for data transfers

FDT has been integrated with PhEDEx at the level of download agent

Basic functionality tested, performance depends on storage systems

FDT deployed as part of DYNES: **Entry point for ANSE**

Zdenek Maxa (Caltech)
The DYNES project targeted extending a production service in Internet2 (and ESnet) and bringing it out to the scientists at major research Universities. This was the first coordinated attempt to use circuit creation on demand at this scale and we found a number of issues that need addressing in the production services.

DYNES has deployed equipment to 54 institutions (Universities or Regional Network Providers). retrofitted 29 sites with OpenFlow capable switches (Dell/Force10 S4810s) allowing us to include OpenFlow options in our toolkit.

DYNES ends officially today (July 31, 2013) and we have setup community support lists to allow the various institutions to self-support moving forward.

DYNES will continue to pursue (in a best-effort way) improving the underlying service resiliency and reliability.

Additionally, we all must work with regional network providers to implement (and document) best practices for protecting circuits created between DYNES sites, respecting the SLAs that are in place.
• The ecosystem in which ANSE operates includes projects and initiatives such as
  – DYNES
    • used by ANSE as the underlying circuit infrastructure in the US
    • interconnects 44 campuses, many of which are part of the LHC computing infrastructure
  – LHCONE (for global service provisioning)
    • ANSE goal is to provide the interface between the point-to-point service and the experiments’ software stacks

• It builds on long-term efforts
  – in various major R&E networks
  – organizations like GLIF, OGF (NSI, NML)

• Software-Defined Networking provides powerful, novel capabilities
  – solving problems currently not solvable commercially
  – use cases and applications currently investigated in projects like OLiMPS, as well as within the LHCONE
THANK YOU! QUESTIONS?

Artur.Barczyk@cern.ch