

# ANSE: Advanced Network Services for Experiments

Shawn McKee / Univ. of Michigan

BigPANDA Meeting / UTA

September 4, 2013



# ANSE at the BigPANDA Meeting?!

- Why am I here giving this talk?
- We have similar goals in our two projects; basically inserting networking into our software infrastructure.
- Since neither project has manpower to waste, we need to consider how best to apportion our activities to get the most overall impact.
- So first some details about ANSE...

# ANSE Project Overview

- ANSE is a project funded by NSF's CC-NIE program
  - Two years funding, started in January 2013, ~3 FTEs
- Collaboration of 4 institutes:
  - Caltech (CMS)
  - University of Michigan (ATLAS)
  - Vanderbilt University (CMS)
  - University of Texas at Arlington (ATLAS)
- **Goal:** Enable strategic workflow planning including network capacity as well as CPU and storage as a co-scheduled resource
- **Path Forward:** Integrate advanced network-aware tools with the mainstream production workflows of ATLAS and CMS
- Network provisioning and in-depth monitoring
- **Complex workflows:** a natural match and a challenge for SDN
- **Exploit state of the art progress in high throughput long distance data transport, network monitoring and control**

# ANSE Objectives

- **Deterministic, optimized workflows**
  - Use network resource allocation along with storage and CPU resource allocation in planning data and job placement
  - Use accurate (as much as possible) information about the network to optimize workflows
  - Improve overall throughput and task times to completion
- **Integrate advanced network-aware tools in the mainstream production workflows of ATLAS and CMS**
  - Use tools and deployed installations where they exist
  - Extend functionality of the tools to match experiments' needs
  - Identify and develop tools and interfaces where they are missing
- Build on several years of invested manpower, tools and ideas

# ANSE Network Focus Areas

- **Monitoring:** Allows Reactive Use
  - React to “events” (State Changes) or Situations in the network
  - Throughput Measurements -> Possible Actions:
    1. Raise Alarm and continue
    2. Abort/restart transfers
    3. Choose different source
  - Topology (+ Site & Path performance) Monitoring -> Possible actions:
    1. Influence source selection
    2. Raise alarm (e.g. extreme cases such as site isolation)
- **Network Control:** Allows Pro-active Use
  - Reserve bandwidth dynamically; prioritize transfers, remote access flows, etc.
  - Co-scheduling of CPU, Storage and Network resources
  - Create Custom Topologies -> optimize infrastructure to match operational conditions: **deadlines, work-profiles**, e.g. during LHC running periods vs reconstruction/re-distribution

# ANSE Methodology

- Use agile, managed bandwidth for tasks with levels of priority along with CPU and disk storage allocation.
  - Allows one to define goals for time-to-completion, with reasonable chance of success
  - Allows one to define metrics of success, such as the rate of work completion with reasonable resource use
  - Allows one to define and achieve “consistent” workflow
- Dynamic circuits a natural match  
(as in DYNES for Tier2s and Tier3s)
- Process-Oriented Approach
  - Measure resource usage and job/task progress in real-time
  - If resource use or rate of progress is not as requested/planned, diagnose, analyze and decide if and when task re-planning is needed
- Classes of work: defined by resources required, estimated time to complete, priority, etc.

# Coupling ANSE to LHC Experiments

- **ANSE** is focused on delivering “Advanced Network Services” for the LHC experiments
  - In ATLAS we are targeting PANDA
  - In CMS we are targeting PhEDEx
  - We have “core” ATLAS/PANDA and CMS/PhEDEx participants
- **ANSE will create a library of network monitoring and services useable by both experiments software stacks**
  - Need to ensure appropriate “hooks” are in place in PANDA and PhEDEx

# Initial ANSE Work

- ANSE first steps are to gather available network information and filter/transform it as needed to make as reliable and useful as possible
  - Getting network metrics from perfSONAR-PS
  - Using SONAR and FAX data from actual transfers
  - How best to filter/normalize and utilize this data?
- Network data exists but not yet complete, consistent or reliable. Working on improving things here.
- Network metrics will be used to inform decision-making when there are choices between sites and/or paths to make.



# Impacts of Networking Monitoring

- Part of the ANSE project is to evaluate how much improvement is to be gained from utilizing knowledge of how networks paths have been performing.
- How “real-time” do we need networking information to be?
  - Does behavior over the last n-hours correlate with behavior over the next m-hours?
- PANDA has the advantage of knowing its incoming workload and how it is scheduling things
  - Can be used to adjust path metrics based upon plans
- **Need to define the best metrics to evaluate impact**

# Beyond Monitoring

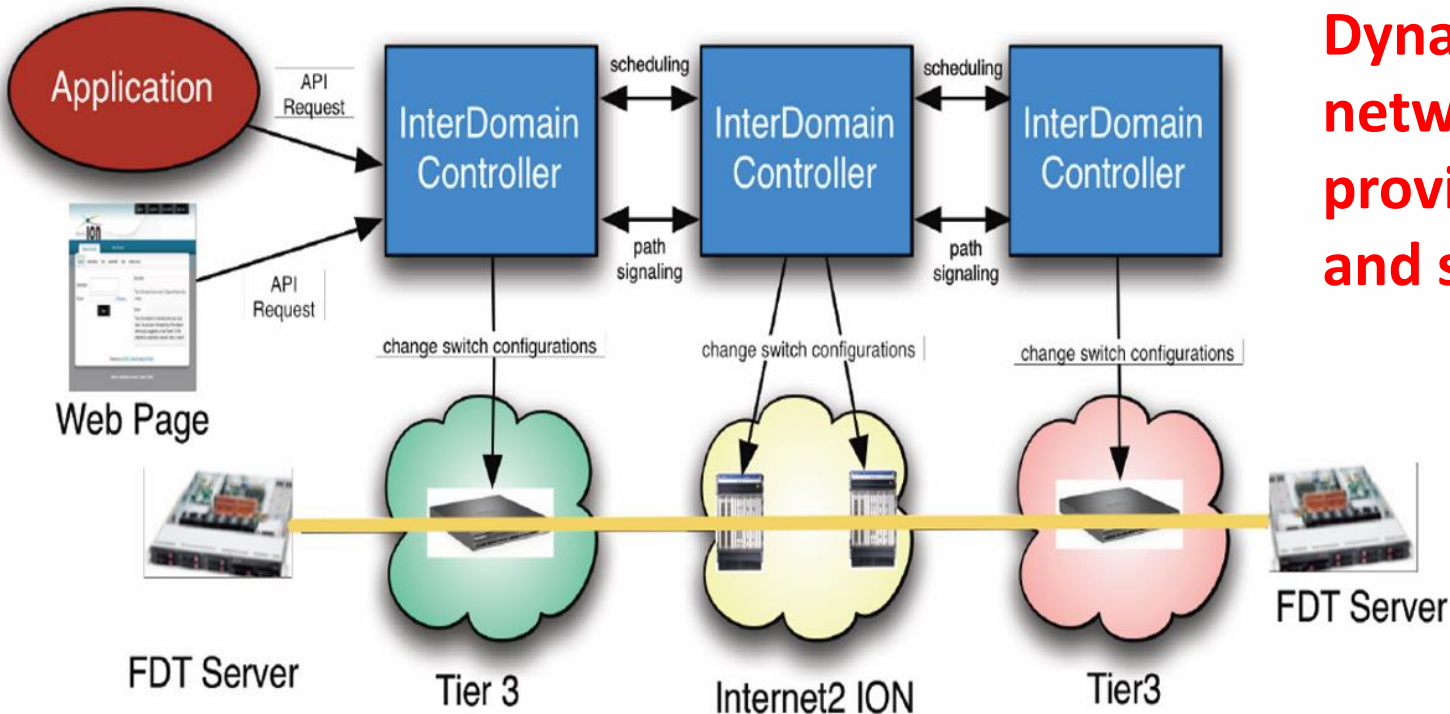
- The consensus is that good monitoring information from the network will help improve our ability to use our resources more effectively but what about negotiating with the network to further improve things?
  - Networks have moved beyond black boxes that transmit bits with varying delay and bandwidth.
  - Users now have the option to negotiate for the service(s) they require.
- Various networking services have been (and are being) developed to better optimize both network resource use and end-user experience:
  - **SDN**: Software Defined Networks; OpenFlow
  - **NSI**: Network Service Interface
  - Dynamic circuits via DYNES/AutoBahn/ION/OSCARS, etc
- **Part of our project is to make sure LHC experiments can utilize and benefit from these developments**

# NSI In a Slide

- “Network Service Interface” is a framework for inter-domain provisioning of connection-oriented services.
  - NSI is an Open Grid Forum (OGF) standard
  - NSI is a consensus standard – open to wide participation and broad based agreement.
  - It is a work in progress. ...and will continue to be evolve and be refined.
- **As a framework, NSI is intended to provide a scalable, secure, and well integrated set of multi-domain architectural elements and service protocols that manage all aspects of services they present:**
- **Connection Reservation and Provisioning:** NSI-CS (version2 – 2012)
- **Topology Exchange** (first edition expected in 2013)
- Others in the wings:
  - Automated Performance Verification
  - Automated Fault Processing
- **Should this be a basis for the ANSE (and BigPANDA) API?**

# What is DYNES?

A distributed virtual cyber-instrument spanning about 40 US universities and 11 Internet2 connectors which interoperates with ESnet, GEANT, APAN, US LHCNet, and many others. Synergetic projects include OliMPS and ANSE.



**Dynamic  
network circuit  
provisioning  
and scheduling**

# DYNES and OpenFlow

- When DYNES started (2010) Software Defined Networking was “alpha” at best
  - We would have liked to start out with something like OpenFlow for use in DYNES
- In our last year we had the opportunity to integrate OpenFlow within DYNES by retrofitting some sites with Dell/Force10 S4810 switches and “beta” OpenFlow firmware.
  - Testing of OpenFlow and the Open Exchange Software Suite (OESS) integration was successful at Michigan (February 2013)
  - Orders for S4810 switches were placed and systems were distributed by July 2013
- This should provide a simpler and more inter-operable DYNES instrument for the future.

# DYNES Status


- **The DYNES project (an NSF MRI 3 year grant) ended July 31...**
- **Last set of sites were included in July 2013**
- **However we are still working on robust circuit creation (issues were discovered in the robustness as scalability of the central circuit services of the R&E backbones)**
- **We have developed a significant infrastructure to provision, deploy and monitor the DYNES instrument that lets us track status**
  - <http://dngs.aglt2.org/DynesNagmap/index.php>
  - <https://dngs.aglt2.org/nagmap/index.php>
- **ANSE depends upon (and assumed) a working DYNES infrastructure. We are trying to make sure we have this in our toolbox.**

# perfSONAR

- Most of you have heard me talk about perfSONAR before (see my ATLAS TIM talk from Tokyo) so I won't repeat those details
- For ANSE (and BigPANDA) we need reliable standard metrics from the network. perfSONAR-PS toolkit installations are an important source we plan to use
  - **However perfSONAR-PS's primary focus is on finding and localizing network issues**
- We are close to having a full WLCG deployment

# WLCG perfSONAR Google Map

← → ↻ <https://grid-deployment.web.cern.ch/grid-deployment/wlcg-ops/perfsonar/conf/monde/V11/> ☆ ⚙ ⌂ 🏠

    **perfSONAR-PS GOOGLE MAPS** Created by: [Anthony HESNAUX](#)  
Managed by: [Simone CAMPANA](#)

Select Cloud Source & Cloud Destination	CLOUD SOURCE <input type="text"/>	CLOUD DESTINATION <input type="text"/>
Select Site Source & Site Destination	SITE SOURCE <input type="text"/>	SITE DESTINATION <input type="text"/>
Select Tier Source & Tier Destination	TIER SOURCE <input type="text"/>	TIER DESTINATION <input type="text"/>

UPDATED AT : [04 - September - 2013\\_13:44](#)



Map data ©2013 MapLink Terms of Use Report a map error



# Incorporating the Network for LHC

- In summary, ANSE goals are to:
  - **Gather** reliable means of estimating network performance between LHC sites.
    - Which metrics? What timescales?
  - **Provide** interfaces for **control** of network characteristics where they exist.
    - Requires “discovery”; AGIS info?
  - Better **enable end-systems to optimize** their ability to utilize the **network connections** available.
    - Some technologies (circuits) need host optimizations
  - **Future proof** our infrastructure by providing generic hooks that can leverage “future” technologies as they are developed.
- **Lots of work and we assume much of this is in-scope for BigPANDA...how to divide the efforts?**

Questions or Comments?

Time for some discussions?

# Additional Slides



# What is OpenFlow?

- OpenFlow is an open standard that enables researchers to run experimental protocols in the campus networks we use every day. OpenFlow is added as a feature to commercial Ethernet switches, routers and wireless access points – and provides a standardized hook to allow researchers to run experiments, without requiring vendors to expose the internal workings of their network devices. OpenFlow is currently being implemented by major vendors, with OpenFlow-enabled switches now commercially available
  - In a classical router or switch, the fast packet forwarding (data path) and the high level routing decisions (control path) occur on the same device. An OpenFlow Switch separates these two functions. The data path portion still resides on the switch, while high-level routing decisions are moved to a separate controller, typically a standard server. The OpenFlow Switch and Controller communicate via the OpenFlow protocol, which defines messages, such as packet-received, send-packet-out, modify-forwarding-table, and get-stats.
  - The data path of an OpenFlow Switch presents a clean flow table abstraction; each flow table entry contains a set of packet fields to match, and an action (such as send-out-port, modify-field, or drop). When an OpenFlow Switch receives a packet it has never seen before, for which it has no matching flow entries, it sends this packet to the controller. The controller then makes a decision on how to handle this packet. It can drop the packet, or it can add a flow entry directing the switch on how to forward similar packets in the future.
- **OpenFlow abstracts the control plane, allowing it to exist centrally in software and giving the network manager programmatic control**