Software Defined Services (SDS) For High Performance
Large Scale Science Data Streams Across 100 Gbps
WANs

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)
Director, StarLight, PI StarLight SDX, Co-PI Chameleon, PI-iGENI, PI-OMNINet (www.startap.net/starlight)

Middleware And Grid Interagency Coordination (MAGIC) – NITRD
Washington DC
February 1, 2017
Macro Network Science Themes

- Transition From Legacy Networks To Networks That Take Full Advantage of IT Architecture and Technology
- Extremely Large Capacity (Multi-Tbps Streams)
- High Degrees of Communication Services Customization
- Highly Programmable Networks
- Network Facilities As Enabling Platforms for Any Type of Service
- Network Virtualization
- Highly Distributed Processes
APIs Based On Messaging and Signaling Protocols
Network Programming Languages
Process Based Virtualization – Multi-Domain Federation –
Policies Cascading Through Architectural Components

Policy Processes
Orchestrator(s)
Northbound Interface
Network OSs
SDN Control Systems
Network Hypervisors
Southbound Interface

State Machines
Westbound Interfaces

Security Processes
Policy Processes
State Data Bases
Mon, Measurements
Real Time Analytics
Eastbound Interfaces

PhyR PhyR PhyR PhyR PhyR VirR VirR VirR VirR
Opendaylight 4th Release: Baryllium

4th Release “Beryllium”
Production-Ready Open SDN Platform
National Science Foundation’s Global Environment for Network Innovations (GENI)

- GENI is funded by the National Science Foundation’s Directorate for Computer and Information Science and Engineering (CISE).
- GENI is a virtual laboratory for exploring future Internets at scale.
- GENI is similar to instruments used by other science disciplines, e.g., Astronomers – Telescopes, HEP - Synchrotrons.
- GENI creates major opportunities to understand, innovate and transform global networks and their interactions with society.
- GENI is dynamic and adaptive.
- GENI opens up new areas of research at the frontiers of network science and engineering, and increases the opportunity for significant socio-economic impact.
Future Cyberinfrastructure

• Large Scale Highly Distributed Infrastructure That Can Support Multiple Empirical Research Testbeds At Scale
• Next Generation GENI, Edge Clouds, IOT, US Ignite, Platform for Advanced Wireless Research (PAWR) and Many Others
• Currently Being Planned – Will Be Designed, Implemented and Operated By Researchers for Researchers
National Science Foundation Global Environment for Network innovations

InstaGENI Racks Network

InstaGENI Network
Oct 2013
International 40G and 100 G ExoGENI Testbed
Chapter: Creating a Worldwide Network For The Global Environment for Network Innovations (GENI) and Related Experimental Environments

R. McGeer, M. Berman, C. Elliott, R. Ricci (Eds.)

The GENI Book

- Provides a foundational overview of GENI’s core architectural concepts
- Presents a detailed discussion of architecture and implementation
- Includes 24 chapters, divided into five sections, which outline GENI from precursors to architecture, development, applications, and then world federation
- Offers an extensive bibliography

This book, edited by four of the leaders of the National Science Foundation’s Global Environment and Network Innovations (GENI) project, gives the reader a tour of the history, architecture, future, and applications of GENI. Built over the past decade by hundreds of leading computer scientists and engineers, GENI is a nationwide network used daily by thousands of computer scientists to explore the next Cloud and Internet and the applications and services they enable, which will transform our communities and our lives. Since by design it runs on existing computing and networking equipment and over the standard commodity Internet, it is poised for explosive growth and transformational impact over the next five years.
The iGENI Initiative Will Design, Develop, Implement, and Operate a Major New National and International Distributed Infrastructure.
• iGENI Will Place the “G” in GENI Making GENI Truly Global.
• iGENI Will Be a Unique Distributed Infrastructure Supporting Research and Development for Next-Generation Network Communication Services and Technologies.
• This Infrastructure Will Be Integrated With Current and Planned GENI Resources, and Operated for Use by GENI Researchers Conducting Experiments that Involve Multiple Aggregates At Multiple Sites.
• iGENI Infrastructure Will Connect Its Resources With Current GENI National Backbone Transport Resources, With Current and Planned GENI Regional Transport Resources, and With International Research Networks and Projects,
The Global Lambda Integrated Facility: a Global Programmable Resource
AutoGOLE Dashboard

Control Plane

Data Plane
NERSC

VLANS:
4012: All hosts
4020: Loop from NERSC to Chicago and back, all NERSC hosts

nsero-tbn-1 NICs:
- 2x40G Mellanox
- 1x40G Chelsio
- 2x10G Myricom
Disk: 24 HDDs

nsero-tbn-2 NICs:
- 4x40G Mellanox
- 1x40G Chelsio
- 1x10G Myricom
Disk: 24 SSDs

NERSC Site Router

10G (MM)

10GE

40GE

exogeni Rack

Alcatel-Lucent 100G SR7750 Router

100G

100G

100G

To Esnet Production Network

Corsa SDN switch to SDN Testbed (coming Summer 2015)

StarLight 100G switch

Note: These hosts have no data disks

star-tbn-1 NICs:
- 4x10G Myricom
- 1x10G Mellanox

star-tbn-2 NICs:
- 4x10G Myricom

star-tbn-3 NICs:
- 4x10G Myricom
- 1x10G Mellanox

Dedicated 100G Network
LHCONE: A global infrastructure for the High Energy Physics (LHC and Belle II) data management
PetaTrans: Petascale Sciences Data Transfer

Singapore
- DTN @NX10G
- SingAREN @100G

Seattle/L.A./SunnyVille
- PacificWave PRP 40G/100G Nodes
- PacificWave /GRPNets @ 2x100G
- PRP Sites

PetaTrans3 @200G

SDXs 100G Switches

GSFC 200G Node

Canarie @300G

Persistence 100G Services Beyond SC16

Washington D.C.
- 100G Switch
- GSFC 200G Node

Amsterdam
- ExoGENI @40G
- ESnet/ANA300 @200G

Montreal
- DTN @100G

Quebec
- ExoGENI @100G

Ottawa
- DTN @40G
ESnet
ENERGY SCIENCES NETWORK

SC16 Scinet, Salt Lake City, UT

6*100 G Scinet

StarLight SDX

All Links 100G unless otherwise noted

<table>
<thead>
<tr>
<th>Demo</th>
<th>VLANs</th>
</tr>
</thead>
<tbody>
<tr>
<td>FNAL demo (ESnet Testbed)</td>
<td>1870 to LBL, 1871 to SL</td>
</tr>
<tr>
<td>NASA demo – Northern path (100G)</td>
<td>1862</td>
</tr>
<tr>
<td>NASA demo – Southern path (100G)</td>
<td>1864</td>
</tr>
<tr>
<td>NRL demo (100G)</td>
<td>VLANs 1849-1849</td>
</tr>
<tr>
<td>Aspera demo (100G loop)</td>
<td>2034</td>
</tr>
<tr>
<td>CalTech/Vanderbilt Demo (80G)</td>
<td>2880</td>
</tr>
<tr>
<td>SLAC Demo (loop)</td>
<td>1700, 1701</td>
</tr>
</tbody>
</table>

SC16 demos – ESnet
Brian Tierney, ESnet 10/28/2016

FILENAME: SC16-Demos-V3.VSD
SCInet TbpsTopology
By Azher Mughal CalTech
Demonstrations of 200 Gbps Disk-to-Disk WAN File Transfers using Parallelism across NVMe Drives

An SC16 Collaborative Initiative Among NASA and Several Partners

SC16 @ Salt Lake City, UT

SCinet NOC Booth #3001

NASA demo @ StarLight/OCC/ICAI R Exhibit Booth #2611

Legend

100-GE

NASA/GSFC-owned

R&D Partners

Max

Dell

Brocade

ESnet

StarLight™

CenturyLink

SCinet

Arista

Intel

Supermicro®

Mellanox

Micro Sata Cables®

Samsung

NASA

HECN

High End Computer Networking
DTN Flows@100 Gbps=>Compute Canada⇔CANARIE⇔StarLight<+>SC16
Programmable Network Measurement of Data Intensive Flows on 100Gbps Networks

Demo1: Programmable Measurement with RESTful APIs
Demo2: Passive & Active Measurement (TCP window size)
Demo3: Passive & Active Measurement (TCP packet loss)

RNC AMIS Team: Yan Luo, PI, University of Massachusetts Lowell; Gabriel Ghinita, Co-PI, University of Massachusetts Boston; Cody Bumgardner, Co-PI, University of Kentucky; Michael McGarry, Co-PI, University of Texas El Paso. Contact: Yan_Luo@uml.edu

Collaborators: Jeo Mambretti, Jim Chen and Fei Yeh, StarLight/iCAIR/Northwestern University; Jeronimo Bezerra, AMPATH/Florida International University
KREONet2 and GLORIAD-KR
KISTI Daejeon ↔ 100 G ↔ StarLight

Soon: Daejeon SDX ↔ StarLight SDX
SINET5 will be a nationwide 100-Gbps backbone network using 100-Gigabit Ethernet technology and connect each pair of nodes with a minimized latency.

**SINET4**
- Star-like topology
- Resource-consuming secondary circuits

**SINET5**
- Fully-meshed topology with redundancy
- Non-resource-consuming secondary paths
A*STAR Singapore

- Singapore Supercomputing Center
  DTN ↔ SingAREN ↔ PacWave ↔ GRPnet ↔ StarLight DTN ↔ SC16
- 50-60 Gbps
Beyond Today’s Internet
Experiencing a Smart Future

Prototype SDX Bioinformatics Exchange: Demonstrating an Essential Use-Case for Personalized Medicine

Robert Grossman, Piers Nash, Allison Heath, Renuka Arya
University of Chicago

Joe Mambretti, Jim Chen
Northwestern University
Biomedical Data Commons:
Flow Orchestration: Control Plane + Data Plane
An Experimental Testbed For Computer Science Research

CHAMELEON:
A LARGE-SCALE, RECONFIGURABLE EXPERIMENTAL ENVIRONMENT FOR CLOUD RESEARCH

Principal Investigator: Kate Keahey

Another SDX Opportunity and Model For “Tenant” Networking – Experimenters As Tenants
CHAMELEON HARDWARE

Switch
Standard Cloud Unit
42 compute
4 storage x2

Core Services
Front End and Data Mover Nodes

Chameleon Core Network
100Gbps uplink public network (each site)

Switch
Standard Cloud Unit
42 compute
4 storage x10

Core Services
3.6 PB Central File Systems, Front End and Data Movers

Heterogeneous Cloud Units
Alternate Processors and Networks

To UTSA, GENI, Future Partners

504 x86 Compute Servers
48 Dist. Storage Servers
102 Heterogeneous Servers
16 Mgmt and Storage Nodes

Chicago
Austin
IMPLEMENTING THE EXPERIMENTAL WORKFLOW

discover resources
- Fine-grained
- Complete
- Up-to-date
- Versioned
- Verifiable

provision resources
- Advance reservations & on-demand
- Fine-grained allocations
- Isolation

configure and interact
- Bare metal
- Deeply reconfigurable
- Map multiple appliances to a lease
- Snapshotting
- Complex Appliances

monitor
- Hardware metrics
- Fine-grained information
- Aggregate and archive

www.chameleoncloud.org
CHI: OVERALL ARCHITECTURE

Portal

Resource discovery

Custom development

Identity Management

Reservation Service (Blazar)

OpenStack

Horizon

Keystone

Glance

Nova

Ironic

Ceilometer

Neutron

special sauce
The Chameleon Network Control API Is the OpenStack Network API -- Neutron, A Cloud Networking Controller And a Networking-as-a-Service (NaaS) Platform.

The Implementation Includes Chameleon Enhancements, e.g., Pre-Set Network Configurations that Experimenters Can Select, Implement, and Toggle Among.

Options The Control Plane Can Support Include Standard Routing Services, Pre-configured VLANs, Extensible VXLANs, Generic Routing Encapsulation (GRE), and OpenFlow with Hybrid Networking.

Because of the Address Limitations of VLANs (~4k), the VXLAN Encapsulation Protocol Architecture Is Important To Multi-Tenant Cloud Providers, Especially For SDN, Because It Extends the Address Space To 16 million With a 24-bit Segment ID, Enabling Provisioning of Large Numbers of Overlay Networks On Shared Infrastructure.
Open Stack SW Architecture For Neutron Reference Platform
Neutron

- Neutron Provides APIs (Via Dashboard Web-Based GUI), interfaces to Devices (e.g., Routers, Switches, virtual routers, virtual switches, and SDN Controllers), And Policy Based Control Software Components.

- Neutron Enables Networks To Be Created and Managed Within IaaS Platforms, e.g., L2 Paths, Routed L3 Paths, IP address Management Processes, and Gateways Through Which It is Possible To Interconnect With External Networks.

- Minimal Neutron Implementation includes At Least One Controller Node (w/ At Least 1 Network Interface On The Management Plane), One Network Node, and One Compute Node.
For Some Services and Applications, VMs Alone Do Not Provide All Capabilities Required, Especially Services Requiring Exceptionally High Performance.

OpenStack Supports Bare Metal Implementations Through the Ironic Environment, Which Can Be Used To Establish, Configure, and Use Bare Metal Nodes.

This Option Requires a Network That Connects to a Boot Image Server That Can Support a PXE Boot Function Over The Network and Other Services, From Neutron, to Configure Bare Metal Nodes.

The Ironic Process Creates a Flat Network, a Single VLAN, Among Bare Metal Nodes.

Chameleon Is Building On This By Developing Options For Network Isolation For Bare Metal Nodes.
Extending Tenant Networks Through Federation

- Tenant Networks Are Not Only Implemented In Cloud Facilities, And Among Cloud Facilities
- They Must Be Extended To Multiple External Sites – Across Multiple Domains – Across The US and Internationally
- Creating Global Private Networks Enhanced Through Federation
- E.g., GENI Has Been Federated With the Future Internet Research and Experimentation (FIRE), With the Smart Applications On Virtual Environments (SAVI) Testbed, with NSFCloud Testbeds, and the International GENI.
IRNC: RXP: StarLight SDX A Software Defined Networking Exchange for Global Science Research and Education

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)

PI IRNC: RXP: StarLight SDX
Co-PI Tom DeFanti, Research Scientist, (tdefanti@soe.ucsd.edu)
California Institute for Telecommunications and Information Technology (Calit2),
University of California, San Diego
Co-Director, StarLight

Co-PI Maxine Brown, Director, (maxine@uic.edu)
Electronic Visualization Laboratory, University of Illinois at Chicago
Co-Director, StarLight

Jim Chen, Associate Director, International Center for Advanced Internet Research, Northwestern University

National Science Foundation
International Research Network Connections Program
Workshop
Chicago, Illinois
May 15, 2015
Planned US SDX Interoperable Fabric
Will Be Contiguous To the StarLight SDX
Next Step: Global Research Platform
Building on CENIC/Pacific Wave and GLIF
CSTNET – In US Over 10 Gbps GRPnet Between Seattle and StarLight
Global Research Platform

• A Emerging International Fabric
• A Specialized Globally Distributed Platform For Science Discovery and Innovation
• Based On State-Of-the-Art-Clouds
• Interconnected With Computational Grids, Supercomputing Centers, Specialized Instruments, et al
• Also, Based On World-Wide 100 Gbps Networks
• Leveraging Advanced Architectural Concepts, e.g., SDN/SDX/SDI – Science DMZs
• Ref: 1\textsuperscript{st} Demonstrations @ SC15, Austin Texas November 2015
• Subsequent Demonstrations @ SC16 Salt Lake City Utah, November 2016

• \textbf{New=> Global Research Platform 100 Gbps Network (GRPnet) On Private Optical Fiber Between PacificWave and StarLight via the PNWGP}
www.startap.net/starlight

Thanks to the NSF, DOE, DARPA Universities, National Labs, International Partners, and Other Supporters