Pacific Research Platform
the US West Coast Big Data Freeway System

presented by: Cees de Laat
many, but not all, slides from:
Tom DeFanti & Larry Smarr et al.

Pacific Research Platform leadership team: (L-R) Larry Smarr, Camille Crittenden, Philip Papadopoulos, Thomas DeFanti, and Frank Würthwein
Why me?

Was there on sabbatical 😊

This personal website contains logs of my sabbatical activities in the spring of 2015. I went on these sabbaticals to be able to engage the community based on content in the ever more and more managerial stuff I was sucked into at UVa. Also my participation in the SURF scientific advisory board made it necessary to study future directions of cyberinfrastructure. Apart from those I always keep on my toes if the research questions we as a group at UVa are studying, are still novel and valid. Directions change very rapidly in particular field of study in System and Network Engineering.

Goals
- Discuss with peers the 5 and 10 year outlook for Cyber Infrastructure. Budgets are shrinking, networks become mature, Supercomputing and High Throughput data compression is now also done with huge commercial cloud centers, so what is the position of specific mission resources in this wild ocean of public capabilities. What do science community need to do ourselves and what can we just buy in the future from the (cloud) market. What do and what do we not need to do on leadership supercomputing. How do we relate to public cloud. We also do not have scientific water. What do we need to do on data at the central level? Do we need a national dataset including data stewardship, etc.
- Next year we are chairing the eutelco workshop. The themes will be CI directions and funding models on national and European level. During the sabbatical seeking out potential speakers.
- Reflection on our own research questions.

Personal
- Take some holiday
- Have Emelie joining me for a couple of days
- Drive the HWY1
- Visit the ocean
- Go to Chicago theatre, music bars, poetry readings
CC-DNI – DIBBs program

• Data Infrastructure Building Blocks (DIBBs) program.
• The DIBBs program encourages development of robust and shared data-centric cyberinfrastructure capabilities to accelerate interdisciplinary and collaborative research in areas of inquiry stimulated by data.
• Effective solutions will bring together cyberinfrastructure expertise and domain researchers, to ensure that the resulting cyberinfrastructure components address the researchers' data needs.

➡➡➡➡➡ Science Drivers!

➡➡➡➡➡ CI plans!
Science Drivers

Particle Physics Data Analysis
- The Large Hadron Collider (LHC). Run 2 will have ~2x the energy, generating ~10x the data volume of Run 1.

Astronomy and Astrophysics Data Analysis
- Includes two data-intensive telescope surveys that are precursors to the Large Synoptic Survey Telescope (LSST)
  - Intermediate Palomar Transient Factory (iPTF)
  - Dark Energy Spectroscopic Instrument (DESI)
- Galaxy Evolution
  - Southern California Center for Galaxy Evolution (CGE)
  - Assembling Galaxies of Resolved Anatomy (AGORA)
- Gravitational Wave Astronomy
  - The Laser Interferometer Gravitational-Wave Observatory (LIGO)

Biomedical Data Analysis
- Cancer Genomics Hub (CG Hub) and Cancer Genomics Browser
- Microbiome and Integrative ‘Omics
- Integrative Structural Biology
Science Drivers (2)

Earth Sciences Data Analysis
- **Data Analysis and Simulation for Earthquakes and Natural Disasters**
  *Pacific Earthquake Engineering Research Center (PEER)*

- **Climate Modeling**
  *National Center for Atmospheric Research (NCAR)*
  *University Corporation for Atmospheric Research (UCAR)*

- **California/Nevada Regional Climate Data Analysis**
  *California Nevada Climate Applications Program (CNAP)*

- **CO2 Subsurface Modeling**

Scalable Visualization, Virtual Reality, and Ultra-Resolution Video
- **Cultural Heritage Data**
- **Networked Scalable Visualization**
- **Virtual Reality Systems**
- **Ultra-Resolution Video Systems**

*Note: different classes of users needing different CI properties, ideally under their own control!*
Science-DMZ

Border Router

WAN

10G Routed

Clean, High-bandwidth path to/from WAN

10G Virtual Circuit

Site / Campus access to Science DMZ resources

Site / Campus LAN

Enterprise Border Router/Firewall

Science DMZ Switch/Router

Dedicated path for virtual circuit traffic

High performance Data Transfer Node with high-speed storage

Per-service security policy control points

Site/Campus Virtual Circuits

High Latency WAN Path

Low Latency LAN Path

High Latency VC Path

perfSONAR

CENIC

ESnet
Science DMZ Superfecta: Engagement

Data Transfer Node
- High performance
- Configured for data transfer
- Proper tools

Engagement
- Partnerships
- Education & Consulting
- Resources & Knowledgebase

perfSONAR
- Enables fault isolation
- Verify correct operation
- Widely deployed in ESnet and other networks, as well as sites and facilities

Network Architecture

Science DMZ
- Dedicated location for DTN
- Proper security
- Easy to deploy - no need to redesign the whole network
<table>
<thead>
<tr>
<th>Institution</th>
<th>Institution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caltech</td>
<td>UC Berkeley</td>
</tr>
<tr>
<td>CENIC / Pacific Wave</td>
<td>UC Davis</td>
</tr>
<tr>
<td>ESnet / LBNL</td>
<td>UC Irvine</td>
</tr>
<tr>
<td>NASA Ames / NREN</td>
<td>UC Los Angeles</td>
</tr>
<tr>
<td>San Diego State University</td>
<td>UC Riverside</td>
</tr>
<tr>
<td>SDSC</td>
<td>UC San Diego</td>
</tr>
<tr>
<td>Stanford University</td>
<td>UC Santa Cruz</td>
</tr>
<tr>
<td>University of Washington</td>
<td></td>
</tr>
<tr>
<td>USC</td>
<td></td>
</tr>
</tbody>
</table>
Science DMZ Data Transfer Nodes - Optical Network Termination Devices: Linux PCs Optimized for Big Data

- **FIONA – Flash I/O Network Appliance**
  - UCOP Rack-mount build: US$7.7K (Blue), $21K (Red)
  - Intel Xeon E5-1650 v3 Haswell, 2x Intel 2.60GHz Xeon E5-2697 v3 14-Core
  - 1TB RAM (expandable to 16TB)
  - Flash Drives: 4TB (up to 16TB)
  - RAID Drives 0 to 112TB
  - NVIDIA Tesla K80 24GB GPU
  - 10GbE/40GbE Adapter (100GbE)
  - Tested speed 37Gbs disk-to-disk
  - UCSD CC-NIE Prism Award & UCOP
    - Phil Papadopoulos & Tom DeFanti
    - Joe Keefe & John Graham
PRPv0 concentrated on the regional aspects of the problem. There are lots of parts to the research data movement challenge. This experiment mostly looked at the inter-campus piece. Over a 10-week period, lots of network and HPC staff at lots of sites collaborated to

- Build a mesh of perfSONAR instances to instrument the network
- Implement MaDDash -- Measurement and Debugging Dashboard
- Deploy Data Transfer Nodes (DTNs)
- Perform GridFTP file transfers to quantify throughput
- Activate an ad-hoc, partial BGP peering mesh across a fabric of 100G links to demonstrate the potential of networks with burst capacity greater than that of a single DTN
- Identify some specific optimizations needed
- Fix a few problems in pursuit of gathering illustrative data
- Identify anomalies for further investigation
Performance for nodes that are close is better than for nodes that are far away.

Network problems that manifest over a distance may not manifest locally.
John Graham’s Network Results
Moving the CineGrid Exchange 30TB
Limited by many 25 Mbyte 4k frame files, file system, ZFS, sata interfaces, etc.
## Data Transfer Testing SSD to SSD
(Using Caltech’s Fast Data Transfer FDT)

<table>
<thead>
<tr>
<th>streams</th>
<th>from</th>
<th>to</th>
<th>GB</th>
<th>seconds</th>
<th>Mb/s</th>
<th>method</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>16 x 120GB SSD in a 2 vdev raidz1 pool at UCSD, 16 x 120GB SSD in a 2 vdev raidz1 pool at UCLA</td>
<td>8x 800GB PCIe NVME</td>
<td>1600</td>
<td>240</td>
<td>53,333</td>
<td>FDT</td>
</tr>
<tr>
<td>1</td>
<td>16 x 120GB SSD in a 2 vdev raidz1 pool at Stanford</td>
<td>/dev/null</td>
<td>100</td>
<td>54</td>
<td>14,834</td>
<td>GFTP</td>
</tr>
<tr>
<td>1</td>
<td>16 x 120GB SSD in a 2 vdev raidz1 pool at Stanford</td>
<td>16 x 4T SATA in a 2 vdev raidz1 pool with 16GB ZIL 960GB ARC on ssd</td>
<td>200</td>
<td>120</td>
<td>13,333</td>
<td>FDT</td>
</tr>
<tr>
<td>2</td>
<td>16 x 120GB SSD in a 2 vdev raidz1 pool at Stanford</td>
<td>16 x 4T SATA in a 2 vdev raidz1 pool with 16GB ZIL 960GB ARC on ssd</td>
<td>400</td>
<td>240</td>
<td>13,333</td>
<td>FDT</td>
</tr>
<tr>
<td>1</td>
<td>16 x 120GB SSD in a 2 vdev raidz1 pool at Stanford</td>
<td>16 x 4T SATA in a 2 vdev raidz1 pool with 16GB ZIL 960GB ARC on ssd</td>
<td>100</td>
<td>66</td>
<td>12,061</td>
<td>GFTP</td>
</tr>
<tr>
<td>4</td>
<td>4 x 800GB PCIe NVME at Caltech</td>
<td>16 x 1T SATA in a 2 vdev raidz1 pool with 16GB ZIL 960GB ARC on ssd</td>
<td>800</td>
<td>1080</td>
<td>5,926</td>
<td>FDT</td>
</tr>
<tr>
<td>1</td>
<td>1 x 800GB PCIe NVME at Caltech</td>
<td>16 x 1T SATA in a 2 vdev raidz1 pool with 16GB ZIL 960GB ARC on ssd</td>
<td>200</td>
<td>300</td>
<td>5,333</td>
<td>FDT</td>
</tr>
<tr>
<td>1</td>
<td>ENSNET, bnl-diskpt1.es.net, <a href="https://cloud.sdsc.edu/auth/v1.0">https://cloud.sdsc.edu/auth/v1.0</a>, USGS edcftp.cr.usgs.gov</td>
<td>16 x 4T SATA in a 2 vdev raidz1 pool with 16GB ZIL 960GB ARC on ssd</td>
<td>1</td>
<td>8</td>
<td>1,039</td>
<td>WGET, swift.py</td>
</tr>
<tr>
<td></td>
<td>UCSD server with 1 SSD</td>
<td>16 x 4T SATA in a 2 vdev raidz1 pool with 16GB ZIL 960GB ARC on ssd</td>
<td>10.74</td>
<td>113</td>
<td>763</td>
<td>LFTP</td>
</tr>
</tbody>
</table>

---

**CENIC**
CENIC Dashboard

CENIC - CENIC Throughput

Next Gen VR Booth

CENIC - CENIC IPv6 Throughput
Optical Exchange as Black Box

Optical Exchange

Switch
TDM
Store & Forward
DWDM mux/demux

TeraByte Email Service

Optical Cross Connect

Oct 29 2004 GridNets conference
WAVE Cluster—20x40G PRP-connected
The WAVE—2013 ~$400,000
35 Megapixels/eye $0.01/pixel

WAVE VR display 7 high by 5 wide HD panels built for the SME Building, UCSD
125,000 Cores (50 nvidia 780s), 200TF (single precision), 1 Terabit Network

CAD Design by Greg Dawe
The Pacific Research Platform is a project to forward the work of advanced researchers and their access to technical infrastructure, with a vision of connecting all the National Science Foundation Campus Cyberinfrastructure grants (NSF CC-NIE & CC-IIE) to research universities within the region, as well as the Department of Energy (DOE) national labs and the San Diego Supercomputer Center (SDSC).

<table>
<thead>
<tr>
<th>Abstract</th>
<th>Science Drivers</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Pacific Research Platform</td>
<td>Particle Physics</td>
</tr>
<tr>
<td>is a project to forward the work of advanced researchers and their</td>
<td>Astronomy and Astrophysics</td>
</tr>
<tr>
<td>access to technical infrastructure, with a vision of connecting all the</td>
<td>Biomedical</td>
</tr>
<tr>
<td>National Science Foundation</td>
<td>Earth Sciences</td>
</tr>
<tr>
<td>Campus Cyberinfrastructure grants (NSF CC-NIE &amp; CC-IIE) to research</td>
<td>Scalable Visualization, Virtual Reality, and Ultra-Resolution Video</td>
</tr>
<tr>
<td>universities within the region, as well as the Department of Energy</td>
<td></td>
</tr>
<tr>
<td>(DOE) national labs and the San Diego Supercomputer Center (SDSC).</td>
<td></td>
</tr>
</tbody>
</table>

US NSF ACI-1540112 and ACI-1541349
Build upon Pacific Wave as a backplane for data-intensive science

- High performance data movement provides capabilities that are otherwise unavailable to scientists
- Integrating Science DMZs across the West Coast
- This capability is extensible, both regionally and nationally
- SDSC’s Andrea Zonca is adapting CILogon for secure access

Goal: scientists can get the data they need, where they need it, when they need it

- PRPv0: a proof of concept experiment to develop and inform requirements for future work.
- Engage with scientists to map their research on to the Pacific Research Platform
Next Steps and Near-term Goals

- Migrate from experiment to persistent infrastructure as part CalREN HPR 100G Layer 3 upgrade; becomes PRPv1
- Sending out the $7,700 FIONA/DTNs & $2,300 PerfSONARs to all 10 UC campuses in September with UC funds; more to non-UC partners in October after NSF contract starts
  - Working with CENIC to identify the people and the connections
  - Upgrading the early FIONAs from March to the new PRPv1 standard
  - Collecting base-line tests State-wide, to Seattle, Chicago, and Amsterdam
  - Identifying and installing security mechanisms
- The PRP will be rolled out in two phases.
  - First, the PRPv1 platform will focus on deploying/maximizing data-sharing to include all member campuses and their science apps
  - Second, the PRPv2 will be rolled out as an advanced, IPv6-based version with robust security and software-defined networking (SDN) features
- Work with campus IT organizations to make “last mile” connections between researchers and the Science DMZ
- PRP Workshop, October 14-16; 60 people signed up/session
- Demos at SC’15, tbd.
Longer-term Goals

- An Integrated West Coast Science DMZ for Data-Intensive Research
- Science DMZ interoperability / integration across regions, nationally, and internationally
- SDN/SDX, …
- Commercial services – Amazon AWS, Microsoft Azure, Google, …
Pacific Research Platform: A Regional Science DMZ

Note: this diagram represents a subset of sites and connections. v1.13 – 20150904
Institutions with Active Archaeology Programs, a science application area of the PRP

Note: This diagram represents a subset of sites and connections.
SARNET
Security Autonomous Response with programmable NETworks

Cyber Security program
PI: CdL
Co-Pi’s: RM, LG, RW
- 400 + 285 + 300 kEuro:
- 3 PhD’s
- Prog & Eng manpower

• Network virtualizations and SDN
• Reasoning
• Risk evaluation
• Trust groups
• Execute response & adaptation
Service Provider Group framework

A Service Provider Group (SPG) is an organisation structure providing a defined service only available if its members collaborate.

Examples:
Service Provider Group Characteristics

- **Autonomous members** acting together on a decision to provide a service none could provide on its own.
- Appears as a *single provider* to a customer.
- Appears as a *collaborative group* to members with standards, rules and policies that are defined, administered, enforced and judged by the group.
- Autonomy in the group: every member signs an agreement *declaring compliance* with common rules, unless local law determines otherwise.
- Membership rules *organizes trust* amongst members and manage group reputation and viability.
Envisioned role of the SPG: define slice archetypes?
Research Questions

• SARNET:
  – Is a cyber security alliance, allowing networks to join/leave freely, feasible?
  – What is needed to organize an alliance, considering the SPG concept?

• Considering future networking concepts:
  – Is a SPG a concept that should identify and arrange slice archetypes e.g. defining cyber-security assurance levels
  – What concerns should the SPG address (e.g. economical-, legal-, administrative-, etc. slice ownerships)?

Collaboration welcomed: delaat@uva.nl
Links

- ESnet fasterdata knowledge base
  - [http://fasterdata.es.net/](http://fasterdata.es.net/)
- Science DMZ paper
  - [http://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf](http://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf)
- Science DMZ email list
  - [https://gab.es.net/mailman/listinfo/sciencedmz](https://gab.es.net/mailman/listinfo/sciencedmz)
- perfSONAR
  - [http://fasterdata.es.net/performance-testing/perfsonar/](http://fasterdata.es.net/performance-testing/perfsonar/)
  - [http://www.perfsonar.net](http://www.perfsonar.net)
- Sample Campus & Regional Cyberinfrastructure Plans
  - [http://fasterdata.es.net/campusCIplanning/](http://fasterdata.es.net/campusCIplanning/)
Questions?