Science DMZ Global Considerations

Presenter: Tom DeFanti, Qualcomm Institute, Co-PI

Larry Smarr, Calit2 Director and PI

Ilkay Altintas, Camille Crittenden, Ken Kreutz-Delgado, Phil Papadopoulos, Tajana Rosing, Frank Wuerthwein, Co-PIs

John Graham, Senior Development Engineer

Dima Mishin, Isaac Nealey, Joel Polizzi, Mark Yashar, Programmers

UC San Diego and UC Berkeley

PRP PACIFIC RESEARCH PLATFORM

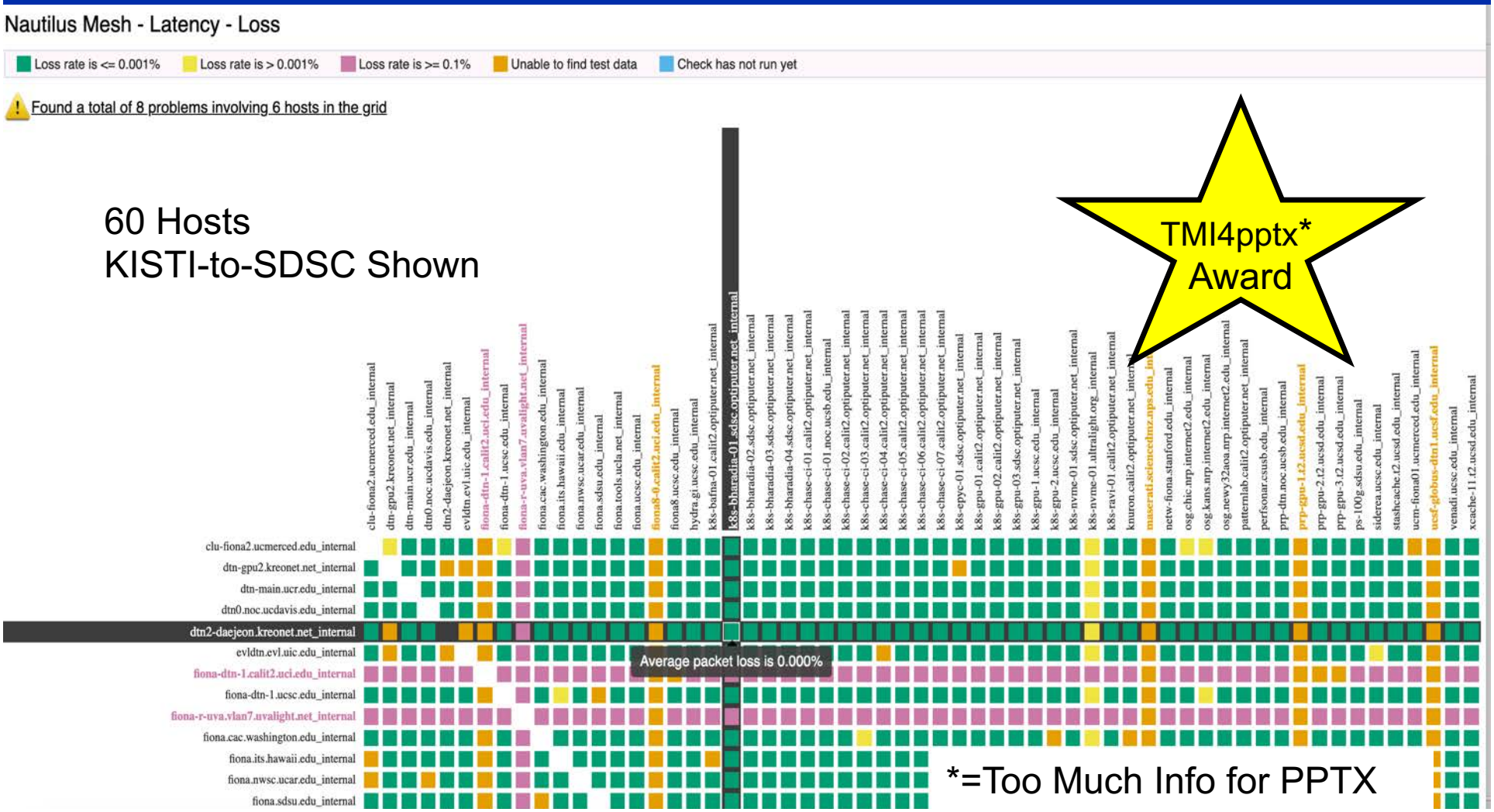
it²

PRP Technical Deliverables
2015 - 2017

- **Phase 0: Tested Layer2 CENIC Networks and FIONAs—early 2015**
- **Phase 1: A Scalable Network for Optimizing Data Transfer—2015-2017**
  - **Layer 3 Data Transfer & Measurement Network**
  - **Tested, Debugged, Measured, Optimized, and MaDDash'ed Layer 3**
  - **Supported Rates up to 9.7/10, 37/40 Gb/s in 10GB Bursts**
  - **Included UvAmsterdam and Korea (KISTI)**
  - **Showed Full Bandwidth Utilization**
  - **Essentially No TCP Backoff on Long Distance Best-Efforts Networks**
- **This is What Most Other Research Platforms are Focusing on—Big Data Transfer**

PRP
PACIFIC RESEARCH PLATFORM

it²


Nautilus Mesh - Latency - Loss
Loss rate is <= 0.001%
Loss rate is > 0.001%
Loss rate is >= 0.1%
Unable to find test data
Check has not run yet
Found a total of 8 problems involving 6 hosts in the grid
60 Hosts
KISTI-to-SDSC Shown
TMI4pptx*
Award
clu-fiona2.ucmerced.edu_internal
dtn-gpu2.kreonet.net_internal
dtn-main.ucr.edu_internal
dtn0.noc.ucdavis.edu_internal
dtn2-daejeon.kreonet.net_internal
evldtn.evl.uic.edu_internal
fiona-dtn-1.calit2.uci.edu_internal
fiona-dtn-1.ucsc.edu_internal
fiona-r-uva.vlan7.uvalight.net_internal
fiona.cac.washington.edu_internal
fiona.its.hawaii.edu_internal
fiona.nwsc.ucar.edu_internal
fiona.sdsu.edu_internal
fiona.tools.ucla.net_internal
fiona.ucsc.edu_internal
fiona8-0.calit2.uci.edu_internal
fiona8.ucsc.edu_internal
hydra.gi.ucsc.edu_internal
k8s-bafna-01.calit2.optiputer.net_internal
k8s-bharadia-01.sdsc.optiputer.net_internal
k8s-bharadia-02.sdsc.optiputer.net_internal
k8s-bharadia-03.sdsc.optiputer.net_internal
k8s-bharadia-04.sdsc.optiputer.net_internal
k8s-chase-ci-01.calit2.optiputer.net_internal
k8s-chase-ci-01.noc.ucsb.edu_internal
k8s-chase-ci-02.calit2.optiputer.net_internal
k8s-chase-ci-03.calit2.optiputer.net_internal
k8s-chase-ci-04.calit2.optiputer.net_internal
k8s-chase-ci-05.calit2.optiputer.net_internal
k8s-chase-ci-06.calit2.optiputer.net_internal
k8s-chase-ci-07.calit2.optiputer.net_internal
k8s-epyc-01.sdsc.optiputer.net_internal
k8s-gpu-01.calit2.optiputer.net_internal
k8s-gpu-02.calit2.optiputer.net_internal
k8s-gpu-03.sdsc.optiputer.net_internal
k8s-gpu-1.ucsc.edu_internal
k8s-gpu-2.ucsc.edu_internal
k8s-nvme-01.sdsc.optiputer.net_internal
k8s-nvme-01.ultralight.org_internal
k8s-ravi-01.calit2.optiputer.net_internal
knuron.calit2.optiputer.net_internal
maserati.sciencedmz.nps.edu_internal
netw-fiona.stanford.edu_internal
osg.chic.nrp.internet2.edu_internal
osg.kans.nrp.internet2.edu_internal
osg.newy32aoa.nrp.internet2.edu_internal
patternlab.calit2.optiputer.net_internal
perfsonar.csusb.edu_internal
prp-dtn.noc.ucsb.edu_internal
prp-gpu-1.t2.ucsd.edu_internal
prp-gpu-2.t2.ucsd.edu_internal
prp-gpu-3.t2.ucsd.edu_internal
ps-100g.sdsu.edu_internal
siderea.ucsc.edu_internal
stashcache.t2.ucsd.edu_internal
ucm-fiona01.ucmerced.edu_internal
ucsf-globus-dtn1.ucsf.edu_internal
venadi.ucsc.edu_internal
xcache-11.t2.ucsd.edu_internal
Average packet loss is 0.000%
*=Too Much Info for PPTX

PRPv1 to PRPv2:
The Transition from Network Diagnosing to Application Support

- **PRPv1 Designed, Built, and Installed ~40 Purpose-built FIONAs, Tuned to Measure and Diagnose End-to-End 10G, 40G and 100G**

- **But, Our PRP NSF Funding Requires Showing Use of the PRP by Scientists and Engineers—It's a Data Grant, not a Networking Grant**

- **Note: Our Scientists Clearly Need More than Bandwidth Tests**
  - **Teams of Scientists Want to Share Their Data at High Speed and Compute on It**
  - **They Need to Interoperate with Commercial and University Clouds**

- **So PRPv2 Added DMZ-Distributed Temporary Storage**
  - **1.7PB total in 14 ~200GB previous PRPv1 FIONAs in Campus DMZs**

PRP
PACIFIC RESEARCH PLATFORM

Why PRPv2 Adopted Kubernetes

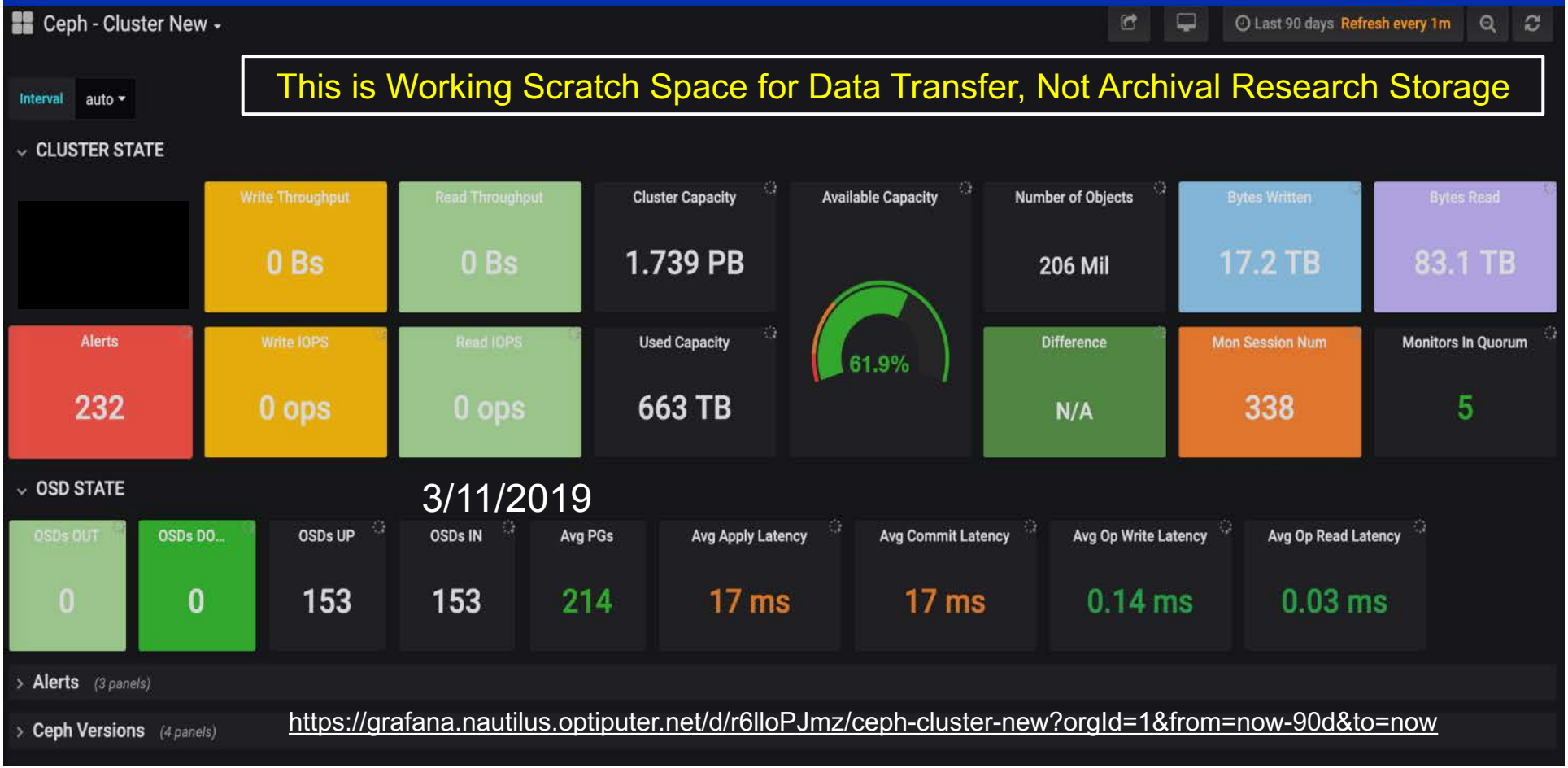
- **PRP FIONAs Are Coupled by Kubernetes Into the "Nautilus Hypercluster"**
  - **Kubernetes "Pods" Encapsulate Application Container(s), Storage Resources, and Execution Options**
  - **Implements PRP Cooperative Research Groups Support with Policy-Based Scheduling by Use of CILogon and Kubernetes Namespaces—704 Users in Namespaces as of 7/15/19**
  - **Allows Cloud Native Storage Integration (e.g., Rook/Ceph/EdgeFS)**
  - **Enables Us to Update Overnight, without local assistance, a RP Scaling Necessity**
  - **Emerging Solutions for Sophisticated SDN Overlay Network, Firewall, and Network Policy Controls**
- **Allows Easy User Job Scaling to Heterogeneous Platforms:**
  - **Deskside, Rack-Mounted, Supercomputers, even IOT Gizmos like ML on Remote Cameras**
  - **Amazon Elastic Container Service for Kubernetes (Amazon EKS)**
  - **Google Kubernetes Engine (GKE) (TensorFlow)**
  - **Microsoft Azure Kubernetes Service (AKS)**

"Kubernetes with Rook/Ceph Allows Us to **Manage** **Petabytes of Distributed Storage and GPUs for Data Science**,
While We Measure and Monitor Network Use."
--John Graham, Calit2/QI UC San Diego

PRP PACIFIC RESEARCH PLATFORM

it²

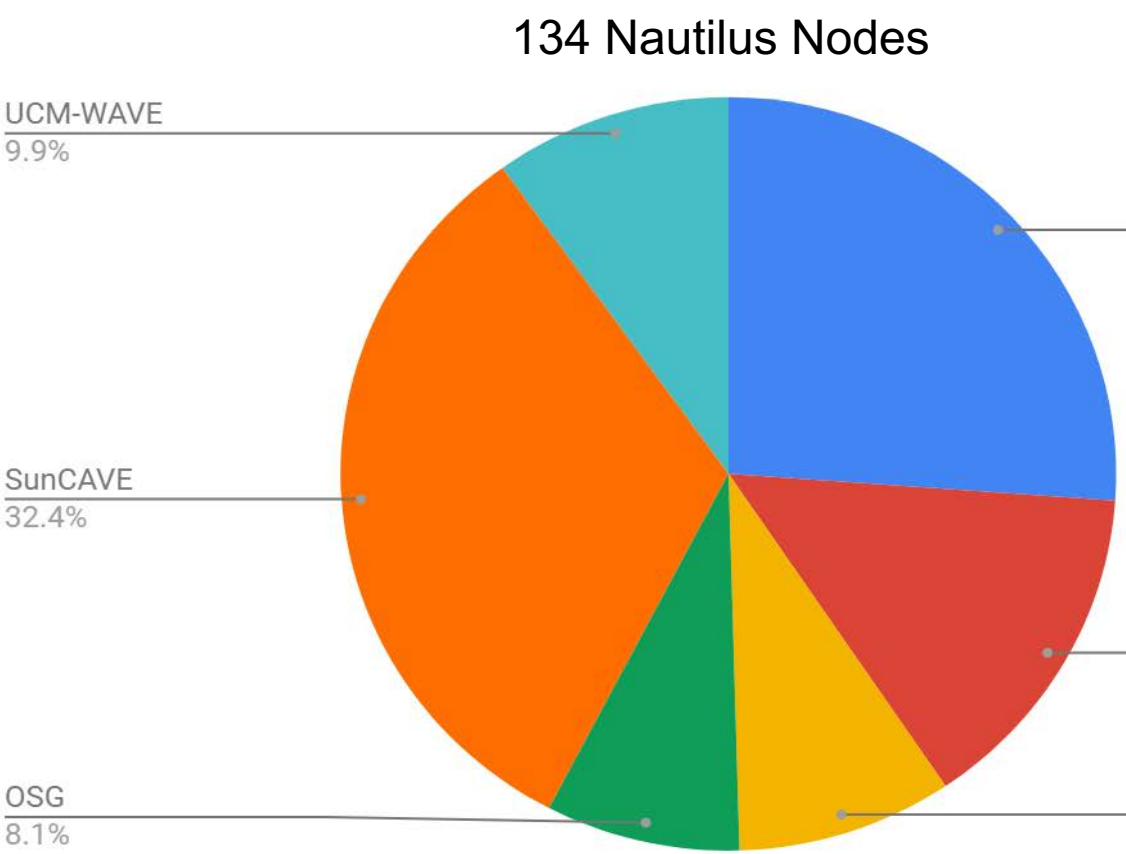
Detailed Real-Time Monitoring of PRP Nautilus:
UCD, UCSD, UCI, UCSB, UCLA, UCR, Stanford, UCAR, UCM, UCSC, UHM Ceph



https://grafana.nautilus.optiputer.net/d/r6lloPJmz/ceph-cluster-new?orgId=1&from=now-90d&to=now

Grafana Showing State of Nautilus 9-10-19

Community Participation Allowed PRP-Paid Nautilus Nodes to be Quadrupled—*Pot Luck Supercomputing*™



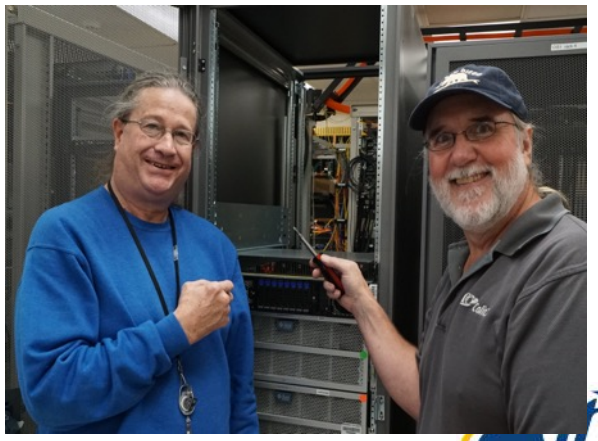
Nautilus

On-Prem for what Commercial Clouds don't do

Road Trip! Installing Community Shared Storage and GPUs in June, December & January at UC Merced, UC Santa Cruz, UC Riverside, and Stanford

160-192TB added to 14 Existing PRPv1 FIONAs

New FIONA8 at UCSC



PACIFIC RESEARCH PLATFORM

2017: PRP Connected 70 UCSD SunCAVE and 20 UCM WAVE 4K Screens to Share VR
2018: Added their 90 Game GPUs to PRP/OSG for Machine Learning Computations

UC Merced WAVE 20 Screens 20 GPUs

UCSD SunCAVE 70 Screens 70 GPUs

PRP PACIFIC RESEARCH PLATFORM

Leveraging UCM Campus Funds and NSF CNS-1456638 & CNS-1730158 at UCSD

it²

Top Nautilus GPU users August 2019					
PI	Campus	August 2019 GPU SU	FIONA8 Equivalent	August 2019 CPU SU	August 2019 Mem SU
Frank Wuerthwein	UCSD	80084	13.90	398124.41	8.13864E+14
Mark Alber	UCR	40761	7.08	37131.21	6.60061E+13
Hao Su	UCSD	16396	2.85	42547.91	2.78718E+14
Nuno Vasconcelos	UCSD	10991	1.91	11218.07	9.11693E+13
Jeff Krichmar	UCI	6587	1.14	6997.06	2.20582E+13
Falko Kuester	UCSD	6211	1.08	35404.91	5.68019E+14
Anshul Kundaje	Stanford	6063	1.05	1481.62	5.38638E+13
Ravi Ramamoorthi	UCSD	4822	0.84	6767.49	3.83436E+13
Larry Smarr	UCSD	4359	0.76	3171.25	2.20892E+13
Manmohan Chandraker	UCSD	3788	0.66	3304.47	1.02188E+14
Tom DeFanti	UCSD	3203	0.56	2040.4	8.82778E+12
Nuno Vasconcelos	UCSD	2293	0.40	3797.22	3.37342E+13
Kurt Schoenhoff	Australia	1921	0.33	4910.91	1.79054E+13
Nuno Vasconcelos	UCSD	1888	0.33	1017.46	1.67571E+13
Dinesh Bharadia	UCSD	1771	0.31	5724.15	2.71821E+13
Padhraic Smyth	UCI	1387	0.24	647.53	1.09787E+13
Jurgen Schulze	UCSD	1330	0.23	10.88	3.9717E+12
Larry Smarr	UCSD	1314	0.23	0.57	2.34185E+12
Jurgen Schulze	UCSD	1306	0.23	0.7	1.92583E+13
Nuno Vasconcelos	UCSD	1209	0.21	5984.29	1.33191E+13
Eric Shearer	UCI	1131	0.20	1308.7	3.85832E+12

Top Nautilus GPU Users in August 2019

FIONA8 equivalent: running an 8-GPU machine 24x7x30

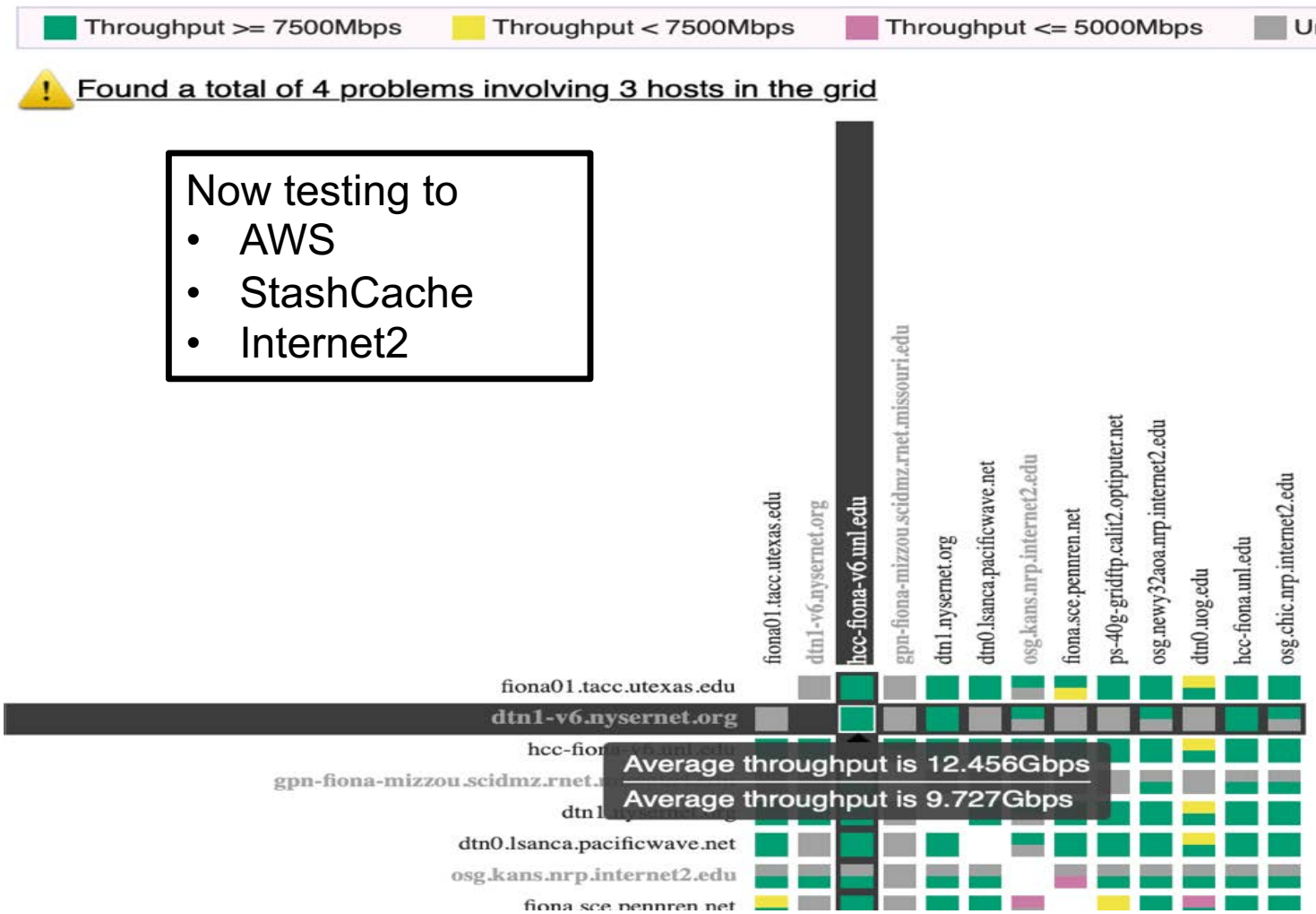
Top User is IceCube in OSG background mode

Others are ML

PRP PACIFIC RESEARCH PLATFORM

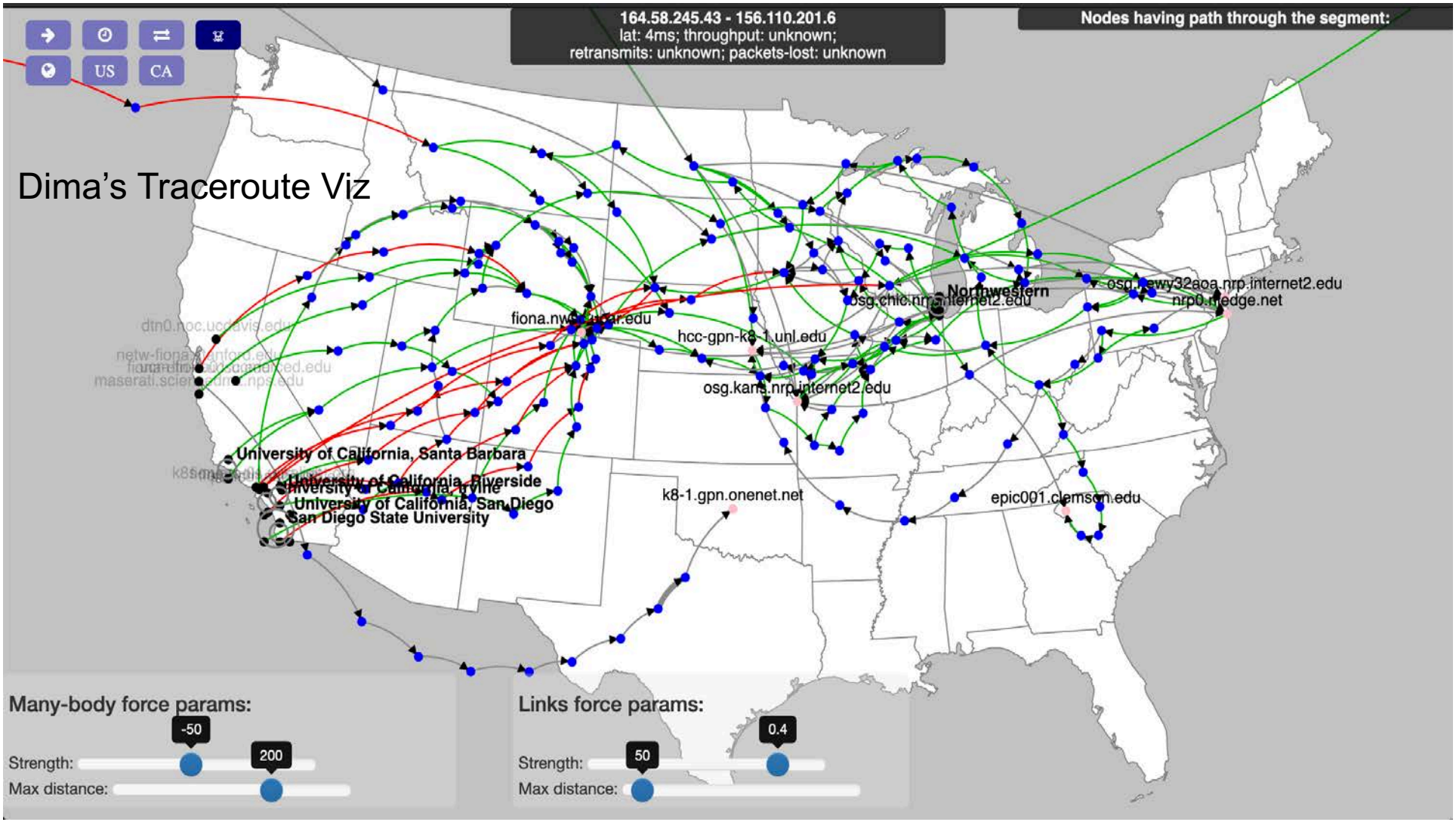
it²

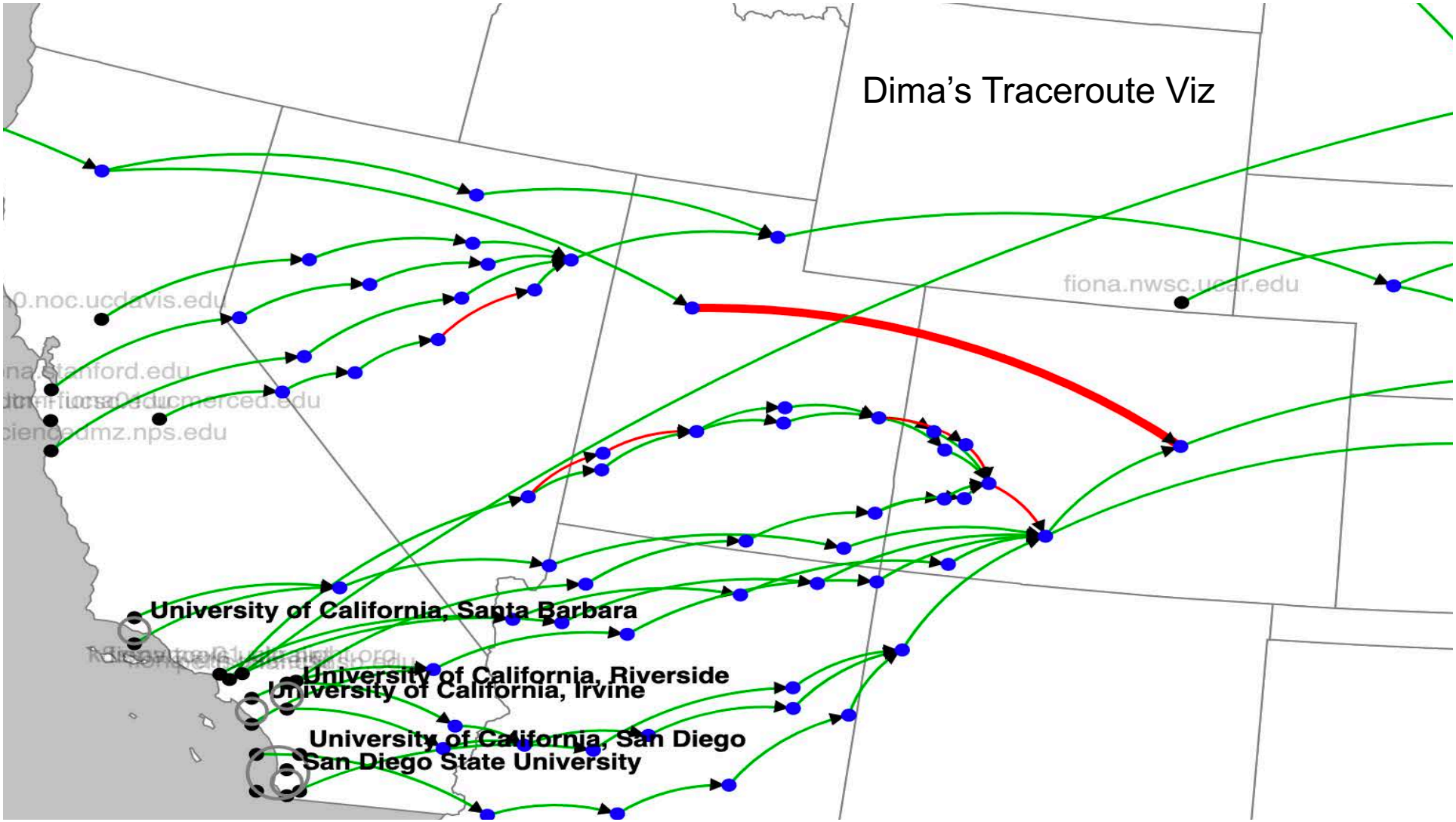
NRP_GridFTP - Throughput

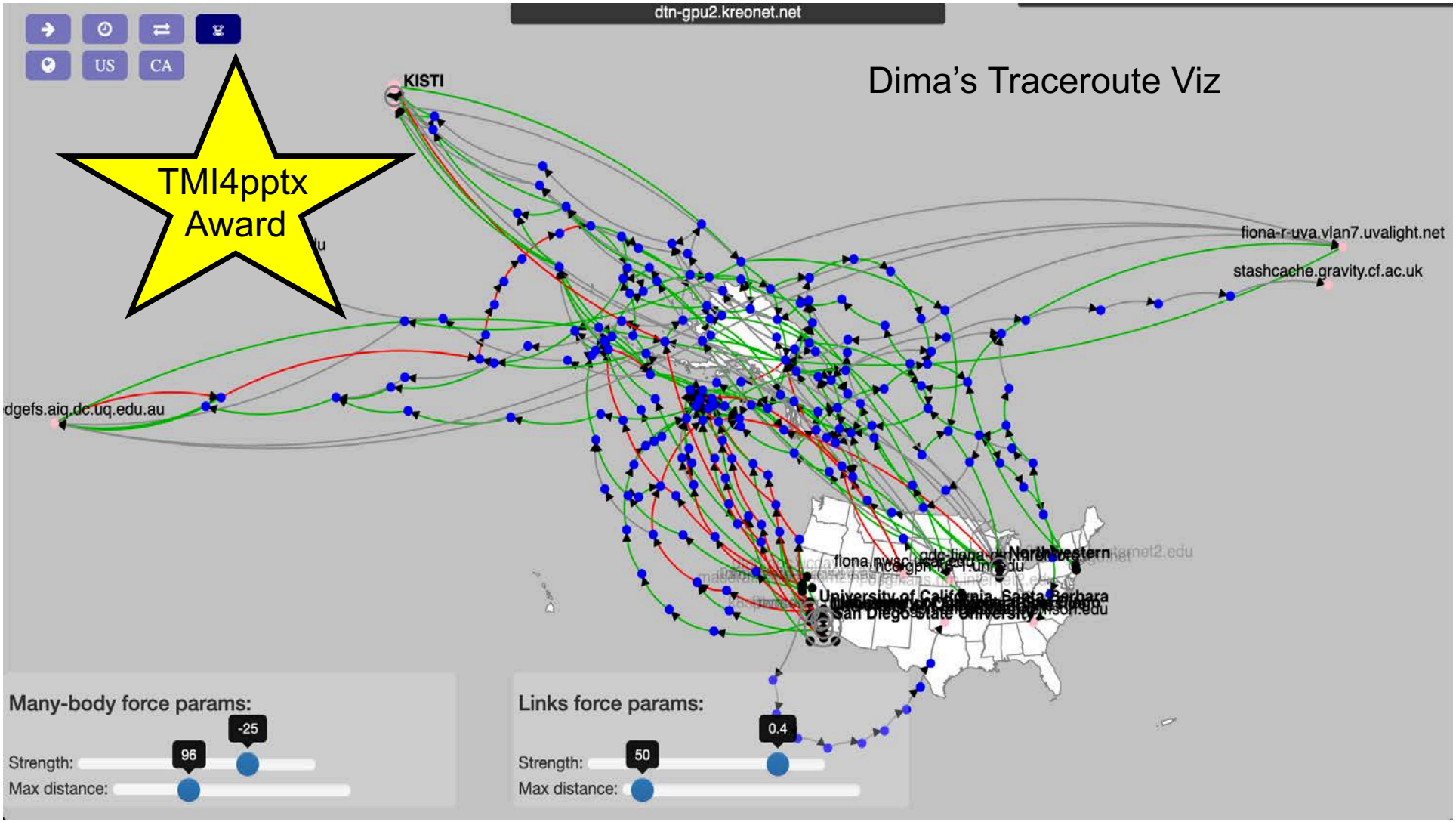


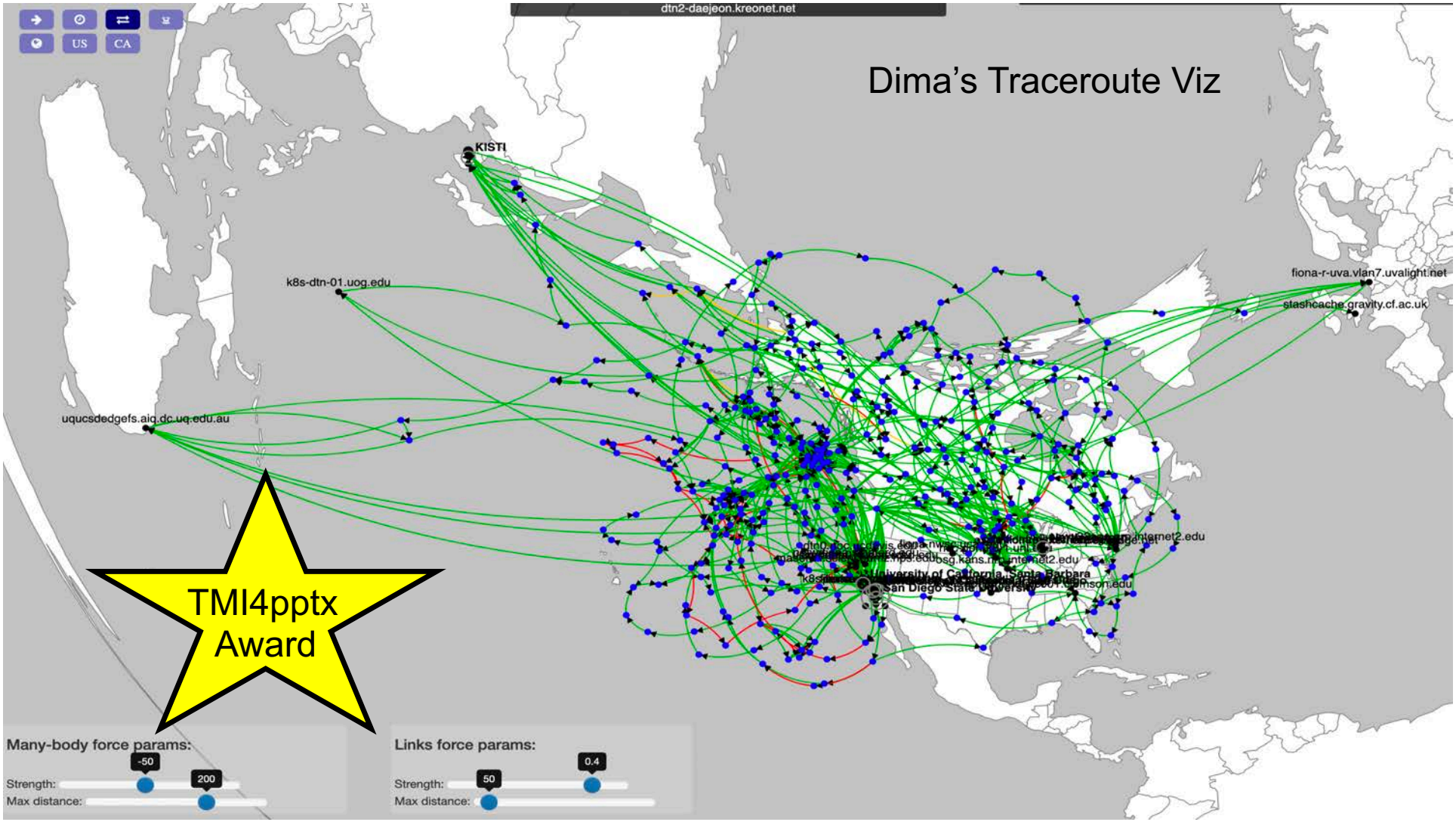
Now testing to

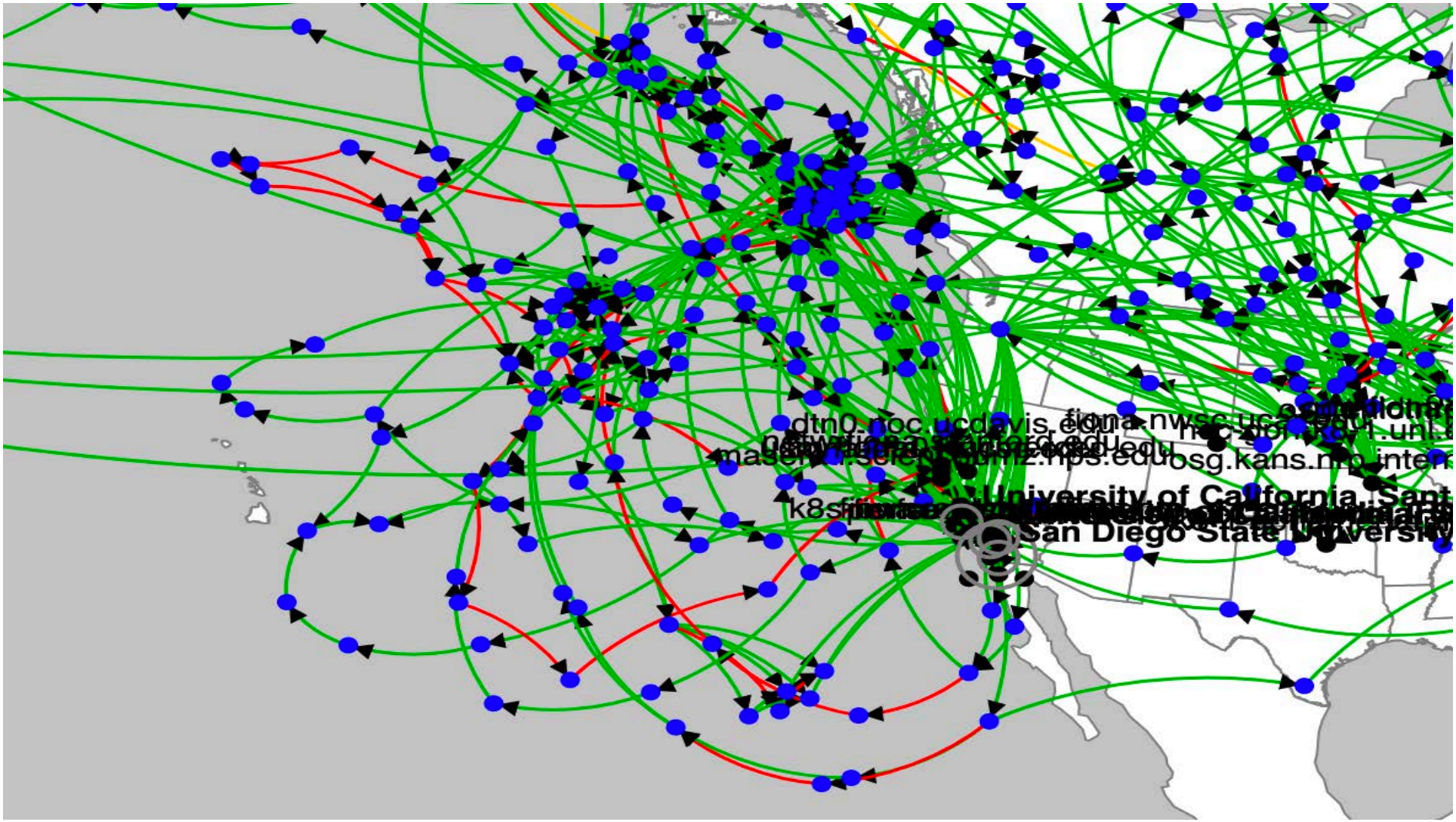
- AWS
- StashCache
- Internet2


Dima’s Traceroute Viz
164.58.245.43 - 156.110.201.6
lat: 4ms; throughput: unknown;
retransmits: unknown; packets-lost: unknown
Nodes having path through the segment:
US
CA
fiona.nwsc.ucar.edu
hcc-gpn-k8-1.unl.edu
osg.kans.nrp.internet2.edu
Northwestern
osg.chic.nrp.internet2.edu
nrp0.mredge.net
University of California, Santa Barbara
University of California, Riverside
University of California, Irvine
University of California, San Diego
San Diego State University
k8-1.gpn.onenet.net
epic001.clemson.edu
Many-body force params:
Strength:
-50
200
Max distance:
Links force params:
Strength:
50
0.4
Max distance:

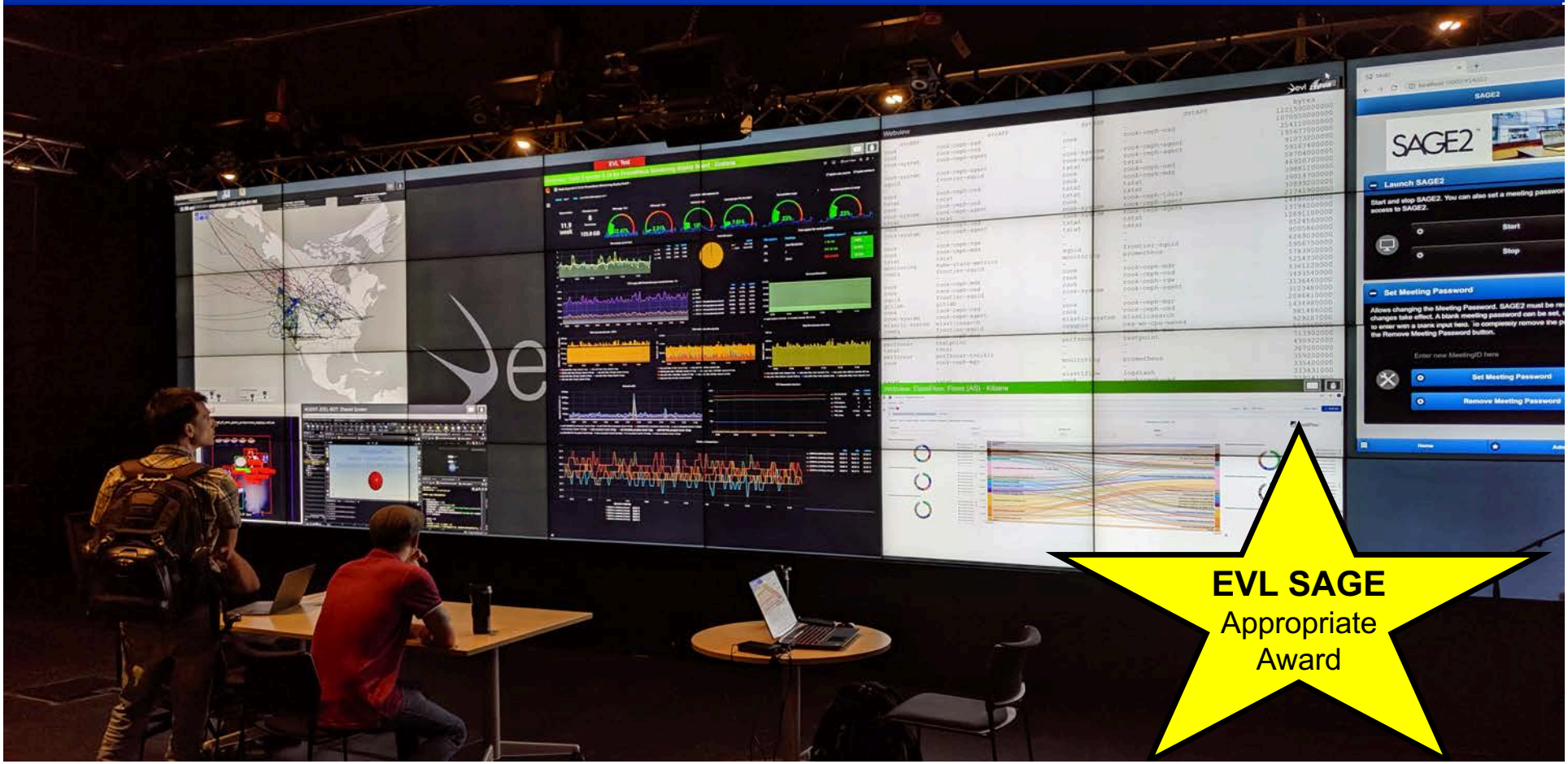

Dima’s Traceroute Viz
fiona.nwsc.ucar.edu
University of California, Santa Barbara
University of California, Riverside
University of California, Irvine
University of California, San Diego
San Diego State University


dtn-gpu2.kreonet.net
US
CA
KISTI
TMI4pptx Award
Dima's Traceroute Viz
fiona-r-uva.vlan7.uvalight.net
stashcache.gravity.cf.ac.uk
dgefs.aiq.dc.uq.edu.au
Many-body force params:
Strength:
96
-25
Max distance:
Links force params:
Strength:
50
0.4
Max distance:

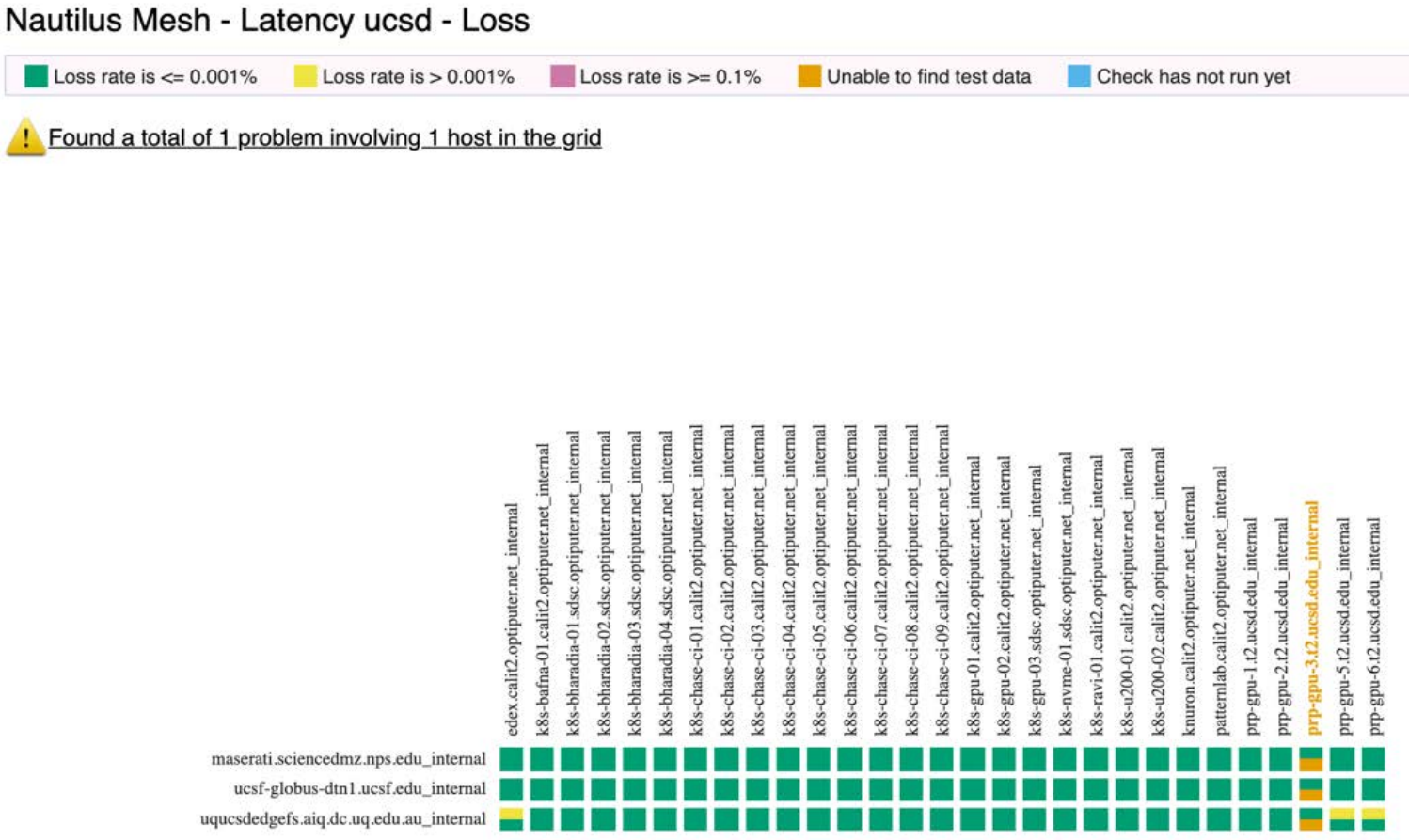

dtn2-daejeon.kreonet.net
US
CA
Dima's Traceroute Viz
KISTI
k8s-dtn-01.uog.edu
uqucsdedgefs.aiq.dc.uq.edu.au
fiona-r-uva.vlan7.uvalight.net
stashcache.gravity.cf.ac.uk
TMI4pptx Award
Many-body force params:
-50
Strength:
200
Max distance:
Links force params:
0.4
Strength:
50
Max distance:


dtn0.noc.ucdavis.edu
University of California, Sant
San Diego State University

NOC of the Future?



Excellent Performance California to UQ



Going Global in Manageable Ways: The Experiment and the Challenge

- **Great Networking with 10-100Gbps Science DMZ Performance is a Necessary *but not Sufficient Condition* to Enable Data-Driven Researchers**
- **They need Science DMZs & DTNs with Low-Cost Storage, Encryption, Large RAM CPUs, GPUs, TPUs, FPGAs, Sensors, and High-Availability Computing**
- **Measuring and Monitoring at all Levels is Key to Better Usage and Security**
- **Compatibility with Google, Microsoft, and Amazon Clouds, and NSF/DOE Supercomputers Helps Ensure Scalability and Continuation—CloudBank**
- **Kubernetes makes it a sane and extensible platform**
- **Open Science Grid and Internet2's NRP Pilot Brings In Global Experience**
- **More Global Partners are Welcome to Join our Pot Luck Supercomputing!**

PRP
PACIFIC RESEARCH PLATFORM

it²

PRP/TNRP/CHASE-CI Support and Community:

- US National Science Foundation (NSF) awards to UCSD, NU, and SDSC
  - CNS-1456638, CNS-1730158, ACI-1540112, ACI-1541349, & OAC-1826967
  - OAC 1450871 (NU) and OAC-1659169 (SDSU)
- UC Office of the President, Calit2 and Calit2's UCSD Qualcomm Institute
- San Diego Supercomputer Center and UCSD's Research IT and Instructional IT
- Partner Campuses: UCB, UCSC, UCI, UCR, UCLA, USC, UCD, UCSB, SDSU, Caltech, NU, UWash UChicago, UIC, UHM, CSUSB, HPWREN, UMo, MSU, NYU, UNeb, UNC,UIUC, UTA/Texas Advanced Computing Center, FIU, KISTI, UVA, AIST
- CENIC, Pacific Wave/PNWGP, StarLight/MREN, The Quilt, Kinber, Great Plains Network, NYSERNet, LEARN, Open Science Grid, Internet2, DOE ESnet, NCAR/UCAR & Wyoming Supercomputing Center, AWS, Google, Microsoft, Cisco