

New network being built to support transfer of big data

March 21 2013

After developing one of the most advanced research communications infrastructures on any university campus over the past decade, the University of California, San Diego, is taking another leap forward in the name of enabling data-intensive science.

The Prism@UCSD project is building a research-defined, end-to-end cyberinfrastructure on the La Jolla [campus](#) capable of supporting bursts of data between facilities that might otherwise cripple the main campus [network](#).

"High-performance cyberinfrastructure is a strategic necessity for a research university," said UC San Diego Chancellor Pradeep K. Khosla. "The Prism network will enable rapid movement of 'Big Data' for multiple, diverse disciplines across campus, including science, engineering, medicine and the arts."

With \$500,000 in funding from the National Science Foundation (NSF), researchers in the UCSD division of the California Institute for Telecommunications and Information Technology (Calit2) are building the network to support researchers in half a dozen data-intensive scientific areas, including genomic sequencing, [climate science](#), electron microscopy, oceanography and physics.

"We've identified a variety of big [data users](#) on this campus who need ten gigabit/s and faster bandwidth to deal with the avalanche of data coming from scientific instruments such as sequencers, microscopes and

computing clusters," said Philip [Papadopoulos](#), principal investigator on the Prism@UCSD project, who splits his time between Calit2 and the university's [San Diego Supercomputer Center](#) (SDSC). "We're starting at 1 Terabit/s of connected capacity through our next-generation modular switch, which is at the center of the Prism network. It can carry 20 times the traffic of our current research network, and it's 100 times the bandwidth of the main campus network."

With the addition of Prism to Calit2's research [network infrastructure](#), the aggregate bandwidth in the Calit2 network will now top one [terabit](#) per second – one trillion bits per second.

"You can think of Prism as the HOV lane," added Papadopoulos, "whereas our very capable campus network represents the slower lanes on the freeway."

"Prism@UCSD is a response to the growing challenge of Big Data," said Calit2 Director Larry Smarr. "The key innovation in Prism@UCSD is to provide end-to-end dedicated large bandwidth to the end-users on campus."

In the past decade, Smarr and Papadopoulos have collaborated on multiple NSF-funded projects to enable cheaper, faster and more energy-efficient scientific computing, storage and visualization. Their OptIPuter project developed a new computer networking paradigm, with optical networks – not computer processors – at the core. That led to Quartzite, an experimental network with reconfigurable optical fiber paths, and wavelength selective switching. The Quartzite core is now six years old, is at full capacity, consumes significant energy, and does not support software-defined networking (SDN) tools such as OpenFlow. Based on those realities and lessons learned in previous projects, Papadopoulos and Smarr were able to create a successful proposal to the National Science Foundation for a more robust, lower energy, faster, and easier to

replicate design.

Prism builds on top of Quartzite, using a next-generation Arista Networks 7405 switch-router, which boasts triple the energy efficiency and four times the capacity of Quartzite's switch. Prism will also expand the existing Calit2-SDSC optical-fiber connection.

"By the time Prism is built out, we will have expanded the SDSC-Calit2 link from 50 to 120Gbps, and it won't cost very much to get it to 160Gbps," said Papadopoulos. "Other campus labs then connect directly to the Prism core at Calit2 with dedicated links of between 20 and 80 Gigabit/s each. The structure allows a Prism-connected lab to saturate any of our external links, no matter where they land on campus. It also enables these labs to share data with each other or utilize high-end resources at SDSC. There is more than enough bandwidth in the switch to accommodate anything you can throw at it." The Arista switch has full bisection bandwidth (as between clusters in a machine room) but it can be deployed at campus scale.

"Prism is the answer to how to move massive volumes of instrument data generated on and off campus to SDSC's powerful Big Data computing and storage resources, Gordon and Data Oasis," said SDSC Director Michael Norman. "Prism will unleash the scientific potential energy of a number of frontier science projects that have been bandwidth limited."

The network will be a hybrid – part "production" infrastructure for real-world use, part "experimental" system for researchers to test out networking ideas. On the production side, the campus is counting on Prism to reduce congestion on the main UCSD network by moving traffic from a few hundred researchers in the most data-intensive fields onto Prism, where they can work with huge data sets that might otherwise clog the campus infrastructure – a state-of-the-art infrastructure that has to serve over 30,000 people.

"The Prism Big Data network also creates a high-capacity 'data freeway' to campus, national or international networks," added Smarr.

Case in point: UCSD physics professor Frank Wuerthwein's lab is the only Open Science Grid (OSG) node on the UCSD campus, and the lab's cluster hosts massive amounts of data from the Large Hadron Collider.

"We want to expand the presence of OSG on this campus," said Wuerthwein, who has signed up to use Prism@UCSD. "For the really big data we are holding – petabytes of Large Hadron Collider data, for instance – it is nice to have a network where we can transmit terabytes of data without killing the campus network in the process."

"The most data-intensive scientific applications get the most value out of using dedicated 'fat' pipes with the ability to accommodate short, extreme-sized bursts of data," said Papadopoulos. "We believe Prism will be the forerunner of specialized, Big Data cyberinfrastructures on many research campuses – and beyond."

Prism will also add a trunk line to the Computer Science and Engineering building, to serve users such as the Center for Networked Systems (CNS). CNS research scientist George Porter and his students use the SEED cluster for Big Data analysis. "One graduate student might work on a 100TB to 200TB data set, and there is only room for one of those at a time on that cluster," said Porter. "So if you wanted to swap data sets, you'd kill the campus network, or you would have to stretch it out over the course of days."

Another major campus user of Prism will be the National Center for Microscopy and Imaging Research (NCMIR), led by professor Mark Ellisman. "We run our own facilities that house petabytes of data distributed across three sites on campus," said Ellisman. "So being able to move around the data to wherever it is needed is extremely important."

We intend to use Prism for our machine room-to-machine room backplane for day-to-day operations."

Added Ellisman: "We will also be able to use it to burst out very large data sets that are generated on NCMIR's array of microscopes and then analyze the data on various Big Data infrastructures that reside physically in different locations on the UCSD campus."

"NCMIR was one of the pioneering science projects that drove the OptIPuter project almost a decade ago," noted Papadopoulos. "It's important for us that a research center with deep knowledge and experience in this arena can really push the envelope and test the limits of how well the Prism network stands up to the needs of the biggest users. Over time, we expect other research groups to follow NCMIR's lead as they begin to handle massive-scale data sets."

According to Papadopoulos, the first constraint in sharing large-scale data at UCSD today is that the many labs that have built up terabytes, cannot easily move the data at will. "This is a first, essential step in a larger data capability that will touch all corners of UCSD and be fundamentally imagined and made real by a very large group of researchers," he noted.

According to Calit2's Smarr, if Prism is a success at UCSD, the project will explore ways to give nearby research labs access to the network – even if they aren't on campus. "UC San Diego has a symbiotic relationship with nearby biotech firms and research institutions on the Torrey Pines Mesa, institutions such as Salk, The Scripps Research Institute, the Sanford Stem Cell Consortium, and Sanford-Burnham," said Smarr. "We are entering the era of integrated, personalized 'omics,' and for San Diego to be a leader, we need to share biomedical data across the Mesa, regardless of which lab generates it."

Most of the NSF funds will be spent on hardware, but Prism will also offer part-time jobs to undergraduate students who help operate the network, while learning about software-defined networking technology. According to Papadopoulos, applicants will have to be "self-starters with a technical bent," preferably with a background in computer science or networking. In addition, a summer workshop aimed at minority-serving institutions will build on Calit2 and SDSC's tradition of diversity outreach.

Provided by University of California - San Diego

Citation: New network being built to support transfer of big data (2013, March 21) retrieved 26 April 2024 from <https://phys.org/news/2013-03-network-built-big.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.