University of California, San Diego

Technologies and Foundations for Robust and Secure Networked Systems

Fall 2009 Newsletter

cns
Center for Networked Systems

Layer 2 Networks: To 100,000 Ports and Beyond

Massive data centers of the future will logically function as single, plug-and-play networks. That's the vision of PortLand — a fault-tolerant, layer 2 data center network fabric capable of scaling to 100,000 nodes and beyond — as outlined by CNS Director Amin Vahdat, senior author of a paper presented in August at SIGCOMM, the premier computer networking conference.

PortLand is fully compatible with existing hardware and routing protocols. It increases a network's inherent scalability, provides baseline support for virtual machines and migration, and dramatically reduces administrative overhead. PortLand also eliminates reliance on a single spanning tree by natively leveraging multipath routing, and it improves fault tolerance.

"With PortLand, we came up with a set of algorithms and protocols that combine the best of layer 2 and layer 3 network fabrics," said CNS's Vahdat, a computer science professor in the Jacobs School of Engineering. "Our goal is to allow data center operators to manage their network as a single fabric. We are working toward a network that administrators can think of as one massive 100,000-port switch seamlessly serving over one million virtual endpoints."

As mega data centers handle more and more of the world's computing and storage needs, data center networking is becoming increasingly important. Loading the front page of any active Facebook user, for example, typically involves over 1,000 servers and takes 300 milliseconds or less.

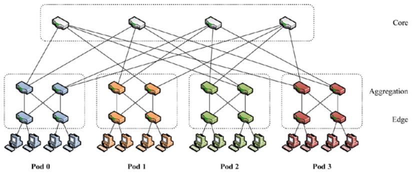
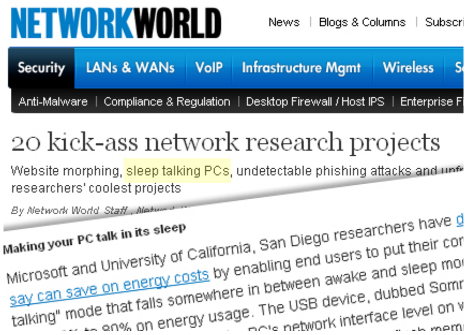


Figure 1: Sample fat tree topology.

"I think PortLand is something that will be useful in the real world. Our goal is to create a network fabric that allows you to buy any commodity server, plug it in and have it just work," said Radhika Niranjan Mysore(pictured above), a UC San Diego computer science graduate student and the first author on the SIGCOMM paper, "Portland: A Scalable Fault-Tolerant Layer 2 Data Center Network Fabric". Other graduate students on the paper include Andreas Pamboris, Nathan Farrington, Nelson Huang,

Continues on page 6

Network World: CNS Work Is "Kick-Ass"


NETWORKWORLD
News | Blogs & Columns | Subscri
Security
LANs & WANs
VoIP
Infrastructure Mgmt
Wireless
Se
Anti-Malware | Compliance & Regulation | Desktop Firewall / Host IPS | Enterprise Fi

20 kick-ass network research projects

Website morphing, sleep talking PCs, undetectable phishing attacks and unfo
researchers' coolest projects

By Network World Staff , Netwo

Making your PC talk in its sleep

Microsoft and University of California, San Diego researchers have d
say can save on energy costs by enabling end users to put their co
talking" mode that falls somewhere in between awake and sleep mo
to 80% on energy usage. The USB device, dubbed Somn
PC's network interface level on w

CNS research presented at last year's SIGCOMM is still winning praise. CNS postdoctoral researcher Kirill Levchenko's work on a new algorithm that could make routers operate more efficiently was designated one of "20 kick-ass network research projects" by Network World magazine in its April 2009 issue. The proposed link-state routing algorithm called Approximate Link state (XL) aims to increase routing efficiency by suppressing updates from parts of the network. Three simple criteria for update propagation are sufficient to guarantee soundness, completeness and bounded optimality for any such algorithm. Published at SIGCOMM 2008, the XL work significantly outperforms standard link-state and distance vector algorithms — in some cases reducing overhead by more than an order of magnitude while having negligible impact on path length.

CNS Members

AT&T · hp invent · Google™ · MOTOROLA · QUALCOMM · NetApp · Sun microsystems · CISCO

In This Issue

Cisco Gift Supports Fight Against Spam

Stefan Savage, Associate Professor in the department of Computer Science and Engineering, received a gift from the Silicon Valley Community Foundation/Cisco University Research Program Fund in support of his project on "Scam Analysis and Defense via Botnet Infiltration."

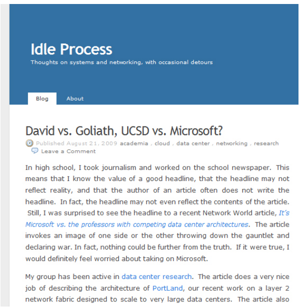
Sun Gift Funds Research on Energy Efficient Design

Tajana Rosing, Assistant Professor in the department of Computer Science and Engineering, received a gift from CNS member Sun Microsystems, Inc. in support of her project on "Energy Efficient Design of Heterogeneous Wireless Sensing Systems for Healthcare Applications."

Rene Cruz Receives 2009 INFOCOM Achievement Award

Rene Cruz (pictured), a professor in the department of Electrical and Computer Engineering, is the recipient of the 2009 INFOCOM achievement award from the IEEE Communications Society. This prestigious award recognizes Professor Cruz`s contributions in the area of communication networks. The award was announced at the 2009 IEEE INFOCOM, IEEE's flagship conference.

CNS Launches New Blog 'Idle Process'



Idle Process

Thoughts on systems and networking, with occasional detours

David vs. Goliath, UCSD vs. Microsoft?

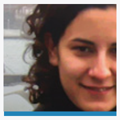
Published August 21, 2009 academia , cloud , data center , networking , research
Leave a Comment

In high school, I took journalism and worked on the school newspaper. This means that I know the value of a good headline, that the headline may not reflect reality, and that the author of an article often does not write the headline. In fact, the headline may not even reflect the contents of the article. Still, I was surprised to see the headline to a recent Network World article, *It's Microsoft vs. the professors with competing data center architectures*. The article invokes an image of one side or the other throwing down the gauntlet and declaring war. In fact, nothing could be further from the truth. If it were true, I would definitely feel worried about taking on Microsoft.

My group has been active in data center research. The article does a very nice job of describing the architecture of PortLand, our recent work on a layer 2 network fabric designed to scale to very large data centers. The article also

CNS Director Amin Vahdat has launched a new blog called *Idle Process*. The blog offers "thoughts on systems and networking, with occasional detours." For everything from CNS news to musings on the trends in such topics as storage infrastructure, cloud computing, data center management, or VLDB (one of those occasional detours), *Idle Process* can be found at: http://idleprocess.wordpress.com/.

After Graduation



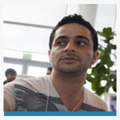
Ayse Coskun, whose advisor is Tajana Rosing, earned a Ph.D. in September 2009 and has accepted a position as an Assistant Professor at Boston University.

Emiran Curtmola accepted a job with Teradata as a software engineer after receiving a Ph.D. in September 2009 (advisor: Alin Deutsch).

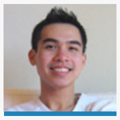
Barath Raghavan graduated with a Ph.D. in May 2009. His advisor was Alex C. Snoeren. Raghavan is now a Visiting Assistant Professor in the Computer Science department at Williams College.

Sebastian Becerra-Licha graduated with an M.S. in June 2009. His advisor was Stefan Savage. Becerra-Licha will be interning at Websense until he joins the Peace Corps in February 2010.

Nelson Huang has graduated with an M.S. and also is a co-author on the SIGCOMM 2009 paper highlighted on the front of this newsletter (with advisor Amin Vahdat). Huang is going to work at Cisco Systems.



Andreas Pamboris has earned his M.S. with Amin Vahdat as his advisor, and is joining the Ph.D. program at Imperial College London.



Pongsakorn Teeraparpwong is going to work for Amazon, armed with an M.S. under Dr. Vahdat.

Two CNS Researchers Receive Prestigious HP Labs Innovation Research Awards

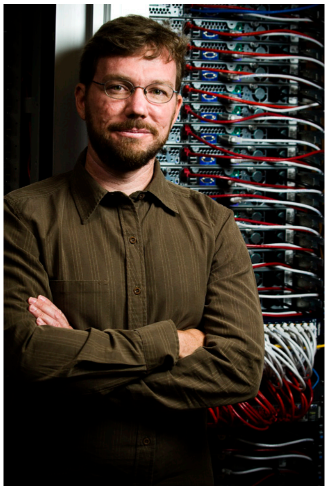
Two computer scientists from the Center for Networked Systems are among 60 professors worldwide to receive awards as part of HP Labs' 2009 Innovation Research Program, which is designed to create opportunities for colleges, universities and research institutes around the world to conduct breakthrough collaborative research with HP. Amin Vahdat and Geoffrey Voelker, professors in UC San Diego's Computer Science and Engineering department, were granted awards as part of this year's competitive open call for proposals. HP, a CNS member company, reviewed nearly 300 proposals from more than 140 universities in 29 countries on a range of topics within the eight high-impact research themes at HP Labs--analytics, cloud, content transformation, digital commercial print, immersive interaction, information management, intelligent infrastructure and sustainability.

"It is an honor for us to have not one, but two of our professors selected by HP Labs for this program that provides critical support for graduate student researchers in their labs," said Keith Marzullo, chair of the Jacobs School's department of Computer Science and Engineering (CSE). "The projects they are pursuing have the potential to do great public good--from squeezing far more bandwidth at lower cost out of large clusters of servers, to delivering a critical blow to spammers who account for the bulk of Internet-based scams."

"Our goal with this program is to collaborate with the brightest minds from around the world to tackle the industry's most complex problems and push the frontiers of fundamental science," said Prith Banerjee, senior vice president, Research, HP and director, HP Labs. "UC San Diego has demonstrated outstanding achievement and a vision that will help inspire technological innovation and address the most complex challenges and opportunities facing the industry in the next decade."

CNS Director **Amin Vahdat** will collaborate with HP Labs on a research initiative focused on interconnecting commodity switches in a fat-tree architecture for clusters consisting of tens of thousands of nodes. "By leveraging strictly commodity switches, we achieve lower cost than existing solutions while simultaneously delivering more bandwidth," said Vahdat, who holds the SAIC Chair in Engineering in the Jacobs School. "We also expect that our approach will be the only way to deliver full bandwidth for large clusters once 10 GigE switches become commodity at the edge, given the current lack of higher-speed Ethernet alternatives at any cost."

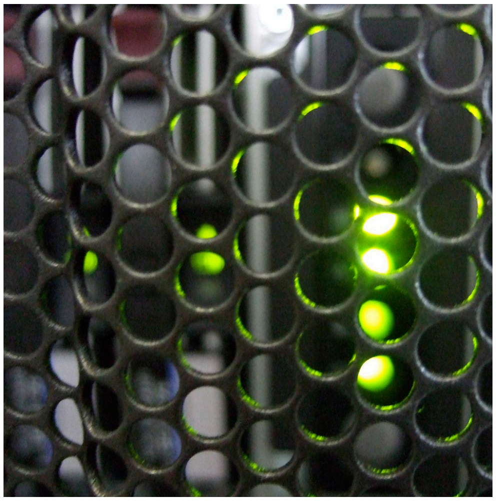
Vahdat's project, "A Scalable, Commodity Data Center Network Architecture," won an HP Labs Innovation Research Award last year. The 2008 award allowed his group to build a hardware-software prototype of a 36-PC scalable data center switch architecture. The 2009 award will allow the researchers to shift the focus from prototype construction to building better scheduling algorithms for dynamically changing communication patterns in the data center. Vahdat and his graduate students also plan to complete specification, validation and implementation of a Location Discovery Protocol--so switches can automatically discover their location in a hierarchical, multi-rooted topology, based only on communication between pairs of locally connected switches and hosts.



Geoffrey Voelker, who is a member of CNS and co-principal investigator on the National Science Foundation-funded Collaborative Center for Internet Epidemiology and Defenses (CCIED), will collaborate with HP Labs on a project titled "Understanding and Exploiting Economic Incentives in Internet-based Scams." According to Voelker, the goal is to better understand the Internet's 'underground economy' and ultimately disrupt its activities. To do so, he said, "we have developed a new technique called 'botnet infiltration' which allows us to measure directly the click-through and conversion rates of Internet spam campaigns in order to get a better understanding of the economics of unsolicited bulk email spam." Voelker, who says a large portion of today's Internet-based crime is fundamentally profit-driven, further explained that "over the last five years the capability of attackers to easily compromise large numbers of Internet hosts has emerged as the backbone of a vibrant criminal economy encompassing spam, phishing, click-fraud, digital denial-of-service (DDoS) extortion and identity theft."

Using messages from 'captured' spam bots, Voelker and co-PI Stefan Savage hope to derive the unique signature of the spammers--who are believed to be relatively few, but who may account for the bulk of spam worldwide. "A small number of organizations likely dominate the market," said Voelker. "Over the next year, our graduate students will help us develop a range of 'fingerprints' to identify different spamming crews while finding ways to undermine their economic models."

CNS Researchers Get GreenLight to Improve Energy Efficiency of Computing



The rapid growth in highly data-intensive scientific research has fueled an explosion in computing facilities and the demand for the electricity to power them. Energy usage per computer server rack will grow from approximately 2 kilowatts (KW) per rack in 2000 to an estimated 30 KW per rack in 2010. Every dollar spent on power for IT equipment requires that another dollar be spent on cooling - equivalent to double the cost of the hardware itself over three years. As a result, cooling and power issues are now becoming a major factor in system design.

To combat this problem, the National Science Foundation is funding UC San Diego's Project GreenLight with $2 million over three years from its Major Research Instrumentation program. The-high profile team of investigators at UC San Diego includes CNS faculty members Amin Vahdat and Tajana Rosing.

GreenLight gets its name from its plan to connect scientists and their labs to more energy-efficient 'green' computer processing and storage systems using photonics — light over optical fiber. GreenLight will be an instrument built to test the energy efficiency of computing systems under real-world conditions, with the ultimate goal of getting computer designers and users in the scientific community to re-think the way they do their jobs. In support of the effort, an additional $600,000 in matching funds is coming from the UCSD division of the California Institute for Telecommunications and Information Technology (Calit2) and the university's Administrative Computing and Telecommunications (ACT) group.

Support also is being provided by CNS member companies Sun Microsystems, Inc. and Cisco Systems, Inc. While Cisco is contributing use of its CiscoWave network on the National LambdaRail (NLR) on an as-available basis, the NSF infrastructure and UCSD matching grants allow UCSD to acquire two Sun Modular Datacenter S20s (Sun MD). One has already been installed, and the second will arrive in Year Three of the project. These are housed in large shipping containers that can accommodate up to 280 servers. To eliminate the need for air conditioning, each Sun MD's closed-loop water-cooling system uses built-in heat exchanges between equipment racks to channel air flow. This allows the unit to cool 25 kilowatts per rack, roughly five times the cooling capacity of typical datacenters. The industry-standard racks can also be placed close together, further reducing the structure's overall eco-footprint and increasing energy efficiency by eliminating dead space.

The GreenLight Instrument will use sensors in the controlled datacenter environment to measure temperature (at 40 points in the air stream), humidity, energy consumption, and other variables, in addition to monitoring the internal measurements of the servers. Researchers hope to use the data to find ways to minimize the power needed to run computers, to make use of novel cooling sources, and to develop software that automates the optimizing of power strategies for each given computing process.

The facility will provide computing and storage services to large-scale projects in five diverse scientific areas: metagenomics; ocean observing; microscopy; bioinformatics; and digital media. Researchers from these fields will be able to carry out quantitative explorations into energy-efficient cyberinfrastructure in a real-world environment.

"We will be running full-scale applications on full-scale computing platforms, so we will be able to draw conclusions about the comparative amount of energy that is consumed by one workload versus another," said Calit2 Director Larry Smarr, who is co-principal investigator on Project GreenLight. "We expect that this new approach will re-define the fundamentals of computer systems engineering and accelerate adoption of a transformative concept for the computer industry — green cyberinfrastructure."

Some of the research groups participating in GreenLight will re-locate servers, switches, computer clusters and related equipment to be deployed inside the first Sun Modular Datacenter. The scientists will continue to operate their equipment virtually and remotely over UCSD's high-performance network, just as if the computers were still in their labs.

Tom DeFanti, Director of Visualization at Calit2 and the Principal Investigator on the project, explained, "The full-scale GreenLight Instrument will measure, monitor and make publicly available real-time sensor outputs using a service-oriented architecture methodology, empowering researchers anywhere to study the energy cost of at-scale scientific computing."

Although the IT industry has begun to develop strategies for 'greening' major corporate data centers, most of the cyberinfrastructure on a university campus involves a complex network of ad hoc and suboptimal energy environments, with clusters placed in small departmental facilities.

According to DeFanti, the project decided to build the GreenLight Instrument around the Sun Modular Datacenter because, "it's the fastest way to construct a controlled experimental facility for energy research purposes." Additionally, the modular structure of the facility also means the GreenLight Instrument can be cloned — unlike bricks-and-mortar computer rooms that cannot be ordered through purchasing.

The computing and systems research will yield new quantitative data to support engineering judgments on comparative "computational work per watt" across full-scale applications running on full-scale computing platforms.

Sun Microsystems research executive
Jud Cooley in front of the GreenLight Instrument

CNS Graduate Students Summer 2009 Internships

Every summer, CNS faculty members coordinate with our corporate partners and affiliates to place CNS graduate students in industrial internships. The CNS Internship Program matches our talented PhD and MS students from the departments of Computer Science and Engineering as well as Electrical and Computer Engineering with mentors in industry who sponsor their work on research projects of common interest. Over the years, these collaborations have resulted in conference papers, journal publications and dissertation topics.

Participants in the CNS 2009 Summer Internship Program:

- Sivashankar Radhakrishnan and Vikram Subramnaya collaborated with Google mentor Tom Everman on "Porting a Wireshark Plugin for an Internal Communication Protocol between Google Switches to Windows."
- Mohammad Al-fares worked with HP Labs mentor Sunjata Banerjee on "Implementation and Evaluation of Datacenter Routing Techniques."
- Eric Rubow collaborated with HP Labs mentors Rick McGeer and Jeff Mogul on "Open Programming Environment for Hardware Switches."
- Dionysios Logothetis worked with Microsoft Research mentors Brad Calder and Vaman Bedekar on the Windows Azure team.
- John McCullough collaborated with Microsoft Research mentors Alec Wolman and John Dunagan on "Stout: A System and API for Building Scalable, Strongly Consistent Internet Applications."
- Radhika Niranjan worked with Microsoft Research mentor John Dunagan on "MicroFragments: Cross Data Center Redundancy Made Cheap."
- Kevin Webb worked with Microsoft Research mentor Emre Kiciman on "Fluxo: A Simple Service Compiler."
- Andreas Pitsillidis collaborated with Microsoft Research mentor Yinglian Xie on "Botnet Spam."
- Todor Ristov worked with Motorola mentor Paul Moroney on "Development and Testing of Cryptographic Libraries and SDKs."
- Edoardo Regini worked with Qualcomm, Inc. mentor Deviprasad Putchala on "Genie."
- Gjergji Zyba collaborated with Thomson Paris Research Lab mentor Christophe Diot on "Pocket Switched Networks."

For information about how to participate in the CNS 2010 Summer Internship Program, please contact CNS Director Amin Vahdat at vahdat@cs.ucsd.edu or Kathryn Krane at kkrane@ucsd.edu.

Continued from page 1...

Pardis Miri, Sivasankar Radhakrishnan and Vikram Subramanya.

Looking for ways to improve data center networking, Vahdat and his team of graduate students revisited the long-standing tradeoffs between layer 2 or Ethernet networks—which route on MAC addresses—and layer 3 networks (which route on IP addresses). Today's data centers are often run on layer 3 networks, but this demands huge numbers of person-hours to set up and maintain the networks. Layer 3 networks also prohibit straightforward implementation of virtual machine migration— limiting flexibility and efforts to reduce energy and cost in the data center.

One of PortLand's key innovations is its location discovery protocol, which opens up the possibility of a scalable layer 2 network. Commodity switches automatically learn their location within the data center topology without any human intervention. These switches, then, assign "Pseudo MAC" (PMAC) addresses to each of the servers they connect to. These PMAC addresses—rather than MAC addresses—are used internally in the network for packet forwarding.

Server behavior remains the same in networks running PortLand. When Server A wants to talk to Server B on the other side of the data center, Server A still sends out an "ARP," which is a request for the MAC address of the computer with which it wants to communicate, based on its IP address. But now, instead of broadcasting this request to the entire network, the switch that received the ARP from Server A talks to a directory service, which returns a PMAC address.

"We have replaced broadcast with a server lookup, and we are forwarding based on PMAC addresses rather than MAC addresses. On the last hop, the switch rewrites the PMAC to be its actual MAC address," said Vahdat.

Added the CNS director: "The students are getting good jobs and internships coming out of this project because they have data center networking skills. Companies are looking for this skill set."

Every Microsecond Counts

Computer scientists have developed an inexpensive solution for diagnosing delays in data center networks as short as tens of millionths of a second — delays that can lead to multimillion-dollar losses for investment banks running automatic stock-trading systems. Similar delays can also slow parallel processing in high performance cluster computing applications run by Fortune 500 companies.

At SIGCOMM, UC San Diego and Purdue University computer scientists presented their invention in a paper titled, "Every Microsecond Counts: Tracking Fine-Grain Latencies with a Lossy Difference Aggregator."

The new approach offers the possibility of diagnosing — at every router within a data center network — packet loss as infrequent as one in a million and delays as brief as tens to microseconds. One microsecond is one millionth of a second. The solution could be implemented in today's router designs at almost zero cost in terms of router hardware and with no performance penalty.

"This is stuff the big traders will be interested in," said George Varghese, a CNS member and computer science professor at the UC San Diego Jacobs School of Engineering. "But more importantly, it's of interest to the router vendors for whom such trading markets are an important vertical."

"This is a diagnostic tool, a potentially extremely important one," added co-author and CNS member Alex Snoeren, a computer science professor in the Jacobs School. "You don't want to just know that you have a network problem, you want to know which router and which application is causing the problem."

If an investment bank's algorithmic stock trading program reacts to information on cheap stocks from an incoming market data feed just 100 microseconds earlier than the competition, it can buy millions of shares and bid up the price of the stock before its competitors' programs can react.

Today's routers are not capable of tracking delays through them at microsecond time scales, so stock exchanges use expensive external boxes to track delays at various key points in the data center network. These boxes, however, are too large and too expensive to be installed at every router in a data center.

"The next step in bringing delay tracking technologies to the routers themselves is to build the hardware implementation; we are looking into that," said first author Ramana Kompella (above), a Purdue computer science professor who earned his Ph.D. in computer science at UC San Diego in 2007.

Simple counters and clever thinking are at the heart of the Lossy Difference Aggregator. The system randomly splits packets coming into a router into groups. Counters add arrival and departure times and then divide by the number of packets for each of the groups separately. As long as the number of packet losses is smaller than the number of packet groups, at least one group will include the information necessary to give a good estimate of the average delay within the router. A series of lightweight counters is the only overhead.

If this invention were built into every router, a data center manager would be able to quickly pinpoint the offending router and interface, explained Kirill Levchenko (pictured below), a UC San Diego postdoctoral researcher who recently earned his Ph.D. in computer science from the university.

Latest in Systems and Networking Innovations Discussed at Summer 2009 Research Review

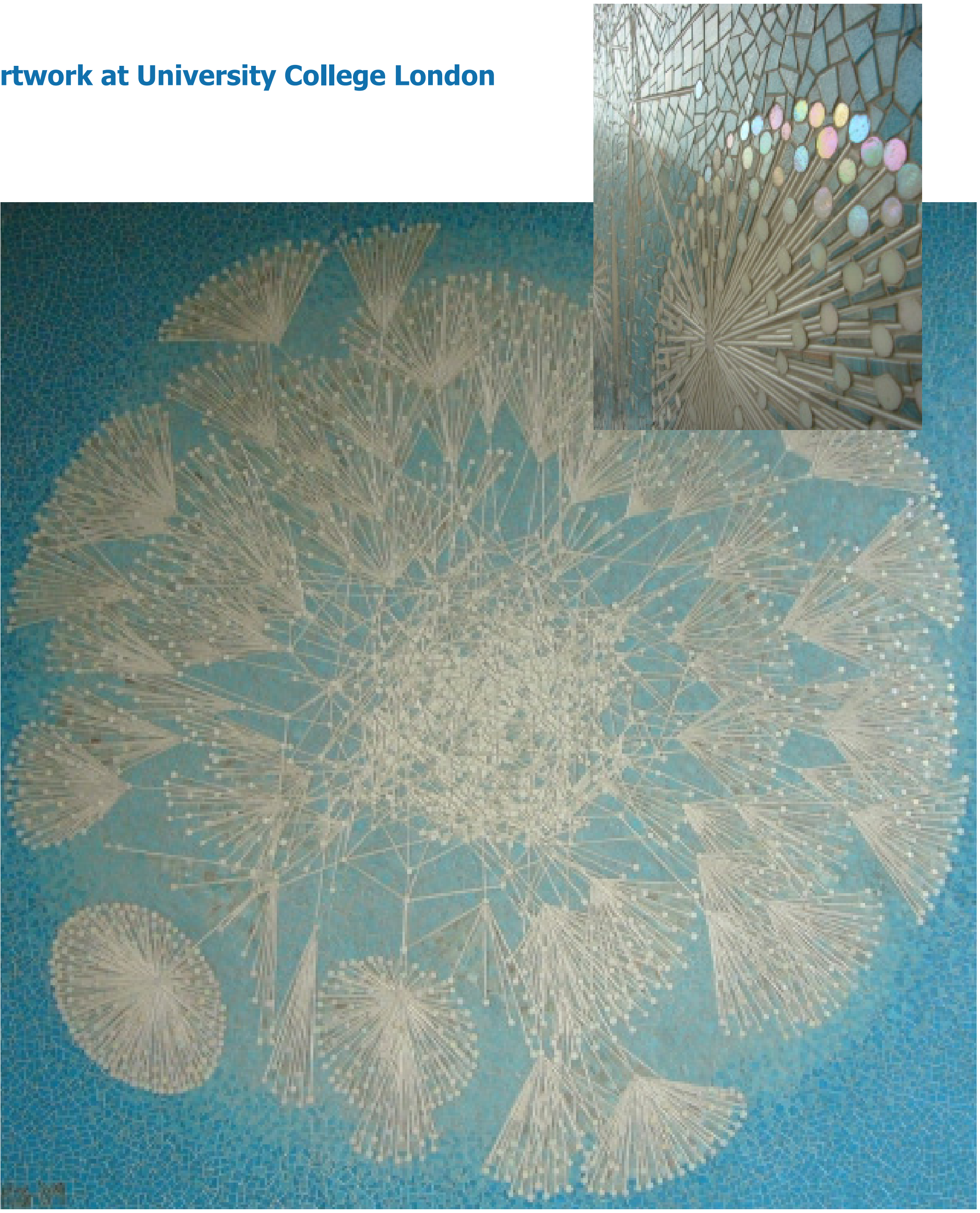
CNS once again broke its attendance record when it held the Summer 2009 CNS Research Review on July 15 and 16 at the University of California, San Diego. The event, which was attended by almost 100 industry guests, faculty, and graduate students, was kicked off with a keynote speech entitled, "Beyond the Microprocessor: Computing's Next Era", by Greg Papadopoulos (at left), Chief Technology Officer and Executive Vice President of Research and Development, Sun Microsystems, Inc. Papadopoulos outlined his vision of the future of the Internet as the "Intercloud," a network of clouds that are unified by a set of protocols and software, and which are segmented into subclouds and "intraclouds" in order to manage local security and predictability. As more devices and processes integrate into the "Intercloud," he said, the inherent limitations of Ethernet technology that are placed upon cloud computing will become more apparent and insupportable. Papadopoulos suggested that current data centers using Ethernet interconnects could be replaced by "microsystem" machines run on large-scale "macrochips." A "macrochip" would be a set of contiguous, optically-interconnected chips enabled by wavelength-division multiplexed silicon-based photonics. By eliminating the flaws of Ethernet interconnects, Papadopoulos suggested that a new wave of innovation in network computing would then be possible involving deep network integration, new memory models and functional accelerators.

Cloud computing, security and energy efficiency once again proved to be prevailing themes as students and faculty proposed projects seeking funds from the CNS research grant program. Research teams also made project summary presentations and delivered final reports on completed grants that had been awarded in July 2007.

Additionally, attendees enjoyed talks by CNS industry representatives. Vidya Narayanan, Principal Engineer at Qualcomm, Inc., presented "Towards a Link Agnostic, Wireless-friendly Peer-to-peer Platform for Applications." Cullen Bash, Principal Research Scientist at HP delivered a talk on "Sustainability and Information Technology;" that explored ways the IT industry could reinvent its current supply chains to create a new ecosystem that drives the reduction of carbon emissions throughout the global economy. Finally, the Technical Director at NetApp, Rory Bolt, discussed virtualization trends and capabilities in his lecture "Storage Considerations for Virtualized Environments."

Another highlight of the Review was the student poster session followed by a dinner reception. Twenty-four students participated in the session which provided them with the opportunity to present their most recent work to our corporate guests and to receive targeted feedback about their research.

CNS Research Inspires Artwork at University College London



A new artwork commissioned by the Computer Science department at University College London proves that the allure of systems and networking can be aesthetic as well as analytical.

The artwork is based upon imagery created by Orbis, a graph generator that employs a number of algorithms to produce random graphs that replicate a given dK-distribution. Two papers by CNS researchers Priya Mahadevan, Dima Krioukov, and Amin Vahdat, appearing in SIGCOMM 2006 and 2007, used Orbis to produce visualizations of Internet topologies. Some of the resulting graphs that modeled a map of the Internet formed what were dubbed “digital dandelions.”

The fragile beauty of the “dandelions” inspired the CS department at University College London to commission an artwork based on the Orbis images for their new building. The artist Hannah Griffiths completed and installed the mosaic Digital Dandelion in 2009.

Mission and Objectives of CNS

The mission of CNS is to develop key technologies and frameworks for networked systems. By combining our research talents and strengths in partnership with industrial leaders, CNS achieves critical mass and relevant focus, accelerating research progress and creating key technologies, frameworks and systems understanding for robust, secure networked systems and innovative new applications. CNS also works to educate the next generation of top students with a perspective on industry-relevant research and to train students on how to continue their leadership throughout their careers. This is accomplished by bringing together leading faculty, students, and companies to investigate the most challenging, interesting and important problems in computer networks.

If you are interested in joining the Center, please contact Director Amin Vahdat at vahdat@cs.ucsd.edu.