COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 13-1 NSF 13-530 04/03/13	FOR NSF USE ONLY
FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S) (Indicate the most specific unit known, i.e. program, division, etc.) ACI - Campus Cyberinfrastrc (CC-NIE)	NSF PROPOSAL NUMBER 1341039

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
04/03/2013	1	05090000 ACI	8080	125084723	09/17/2013 2:14pm S

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)	SHOW PREVIOUS AWARD NO. IF THIS IS	IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)
941539563	☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE University of California-Santa Cruz	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE 1156 High Street SANTA CRUZ, CA 95064-1077
AWARDEE ORGANIZATION CODE (IF KNOWN) 0013219000	
NAME OF PRIMARY PLACE OF PERF University of California-Santa Cruz	ADDRESS OF PRIMARY PLACE OF PERF, INCLUDING 9 DIGIT ZIP CODE University of California-Santa Cruz 1156 High St Santa Cruz ,CA ,950641077 ,US.

IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)
☐ SMALL BUSINESS ☐ MINORITY BUSINESS
☐ FOR-PROFIT ORGANIZATION ☐ WOMAN-OWNED BUSINESS
☐ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT **CC-NIE Networking Infrastructure: 100 Gb/s Science DMZ**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ 500,000	24 months	01/01/14	

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW
☐ BEGINNING INVESTIGATOR (GPG I.G.2)
☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C.1.e)
☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG I.D, II.C.1.d)
☐ HISTORIC PLACES (GPG II.C.2.j)
☐ EAGER* (GPG II.D.2) ☐ RAPID** (GPG II.D.1)
☐ VERTEBRATE ANIMALS (GPG II.D.6) IACUC App. Date ________
PHS Animal Welfare Assurance Number ________
☐ HUMAN SUBJECTS (GPG II.D.7) Human Subjects Assurance Number ________
Exemption Subsection ________ or IRB App. Date ________
☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED (GPG II.C.2.j)
____

PI/PD DEPARTMENT Computer Engineering	PI/PD POSTAL ADDRESS 1156 High Street SANTA CRUZ, CA 950644107 United States
PI/PD FAX NUMBER 831-459-1640	

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Email Address
PI/PD NAME Bradley R Smith	DPhil	2003	831-459-5278	brad@soe.ucsc.edu
CO-PI/PD Mary Doyle	DPhil	1998	831-459-5278	mdoyle1@ucsc.edu
CO-PI/PD				
CO-PI/PD				
CO-PI/PD				

PROJECT SUMMARY

Overview:

Faculty and staff at the University of California at Santa Cruz are engaged in several very large projects that involve managing and analyzing massive amounts of data. For example, the Astrophysics group led by Prof. Stan Woosley generates 30 TB of data each week that they would like to transfer to the local Hyades cluster for further analysis. The Santa Cruz Institute for Particle Physics is involved in a number of network-intensive research projects, including ATLAS, VERITAS, DES, and LSST, that are currently bandwidth-limited by UCSC's existing network connections. The Genome Browser (http://genome.ucsc.edu), which is used by researchers around the world for an interactive, graphical display for comparing genomes, handles peak loads that saturate its current 1 Gb/s link. The Cancer Genome Hub project, sponsored by NIH, anticipates transfers of almost 33 PB of data between the CGHub site and researchers from across the country over the three and a half years of the project. The CGHub data repository itself is at SDSC, but the repository developers and operators reside at UCSC and require high performance access to the data for the purposes of development, analysis, and archiving.

The goal of this project is to redesign the research network at UCSC to provide the research projects with the high performance network support they need. The architecture of the network will be flexible and expandable, so as new projects come online, the network can be adapted to handle their needs. The network architecture will provide enhanced measurement and monitoring capabilities necessary for the support of high performance network environment, as well as the ability to deploy novel and experimental technologies (notably, OpenFlow will enable experiments with routing). Further, requested support personnel will act as a bridge between the research projects using this enhanced network and the central campus IT staff managing the network, so researchers can obtain the optimum network performance that they need, and the network itself can be tuned as appropriate to minimize overhead.

These improvements will enable research projects requiring high performance networking to take full advantage of the throughput of high speed networks, such as the CENIC HPR network, connected to the campus border.

Intellectual Merit :

The intellectual merit of this work lies in the planning of the new network infrastructure and architecture. How does one make the network as flexible as possible without adversely affecting performance? Equally important, how do the specific needs of a research project affect the network architecture and infrastructure? This requires an understanding of the research projects' needs, and the management structure proposed here creates a feedback mechanism to acquire this understanding. Finally, the question of how to bring together diverse constituencies in widely varied fields to ensure needs are met as much as possible is itself a problem of no small intellectual merit.

Broader Impacts :

Given the wide diversity of research projects at UCSC, this project will have broad impact on research conducted in a variety of disciplines. This applies particularly to collaborative projects such as the Genomics and SCIPP projects. Further, the instrumentation supports research into data management itself: for example, how do we optimize performance in a network carrying large amounts of data for many projects? In addition, this project will employ an engineering undergraduate student who will work closely with research project staff dealing with issues of large-scale data movement and campus staff managing the research network. This will not only enhance their education by having them apply what they learn in class, but will also help develop a pool of high performance networking experts familiar with managing the motion of large data sets.

Project Description

1. Introduction

The University of California at Santa Cruz is home to a number of leading research centers in science and engineering. The Center for Biomolecular Science and Engineering supports the UCSC Genome Browser, a crucial resource for the international scientific community, and develops and operates the Cancer Genomics Hub (CGHub), a secure repository for storing, cataloging, and accessing cancer genome sequences, alignments, and mutation information from The Cancer Genome Atlas (TCGA) consortium and related projects.

UCSC is home to one of the world's leading centers for research in computational astrophysics. At any time, about twenty faculty in four departments (Applied Math & Statistics, Astronomy & Astrophysics, Earth & Planetary Sciences, and Physics) and at least fifty graduate students and postdocs are using a variety of on-campus and off-campus computer facilities to address some of the fundamental scientific questions of our time, from the structure of the early Universe and the nature of dark matter, through the formation of galaxies, stars, and planetary systems, to impacts in the solar system. Over longer periods, many more graduate students and postdocs are involved. UCSC Computational Astrophysics papers are highly cited and their simulations are frequently featured in the media.

The Santa Cruz Institute for Particle Physics (SCIPP) is an organized research unit within the University of California system. SCIPP's scientific and technical staff are and have been involved in several cutting edge research projects for more than 25 years. The primary focus is experimental and theoretical particle physics and particle astrophysics, including the development of technologies needed to advance that research. SCIPP is also pursuing the application of those technologies to other scientific fields such as neuroscience and biomedicine. The Institute is recognized as a leader in the development of custom readout electronics and silicon micro-strip sensors for state-of-the-art particle detection systems.

Supporting these activities at UCSC is a campus network that provides many of the features of a science DMZ. Research traffic is routed to avoid the firewalls and intrusion prevention systems that protect the administrative systems of the University. Thus, much less filtering occurs, enabling better performance and more flexibility than in a network where the traffic is routed through the administrative firewalls and other security appliances.

Nevertheless, over the past few years it has become clear that the "big data" produced and used by campus research projects will require a faster, more robust, and more flexible network on campus that is accessible by more researchers. Such a network will require sophisticated monitoring for performance and security purposes, enhancements to a number of buildings on campus where network bottlenecks exist today, and the ability to change routing quickly and easily as traffic characteristics change.

UCSC has a strong track record of collaboration between the campus research community and the Division of Information Technology Services; this proposal rebuilds a portion of the campus network dedicated to research that emphasizes performance and research value by providing an infrastructure which is substantially independent of the general campus network.

While not a requirement of the program, a key element of the UCSC network plan involves building the capabilities identified here to support collaborative project components at other institutions. As we work to address campus network shortcomings through this program, we also continue to work with CENIC and Internet2 to identify issues that affect network performance external to UCSC, including the path between UCSC and SDSC (where CGHub is located), between SDSC and CGHub collaborators across the nation, between campus astrophysics researchers and the national labs where their largest computations are hosted, as well as between campus SCIPP researchers and the various sources of LHC data they consume. Our existing partnership with CENIC in this regard is substantial and includes participation as part of the next-generation network design team within CENIC.

In addition, this project will be tied to the educational programs of the associated researchers. Faculty will be able to integrate research visualizations into their classes, as well as providing access to computational systems and algorithms online. For students in most computer labs on campus or with wired laptops on campus or in residence halls, visualizations consistent with research laboratory quality would become possible.

The plans for integrated monitoring will allow us to obtain metrics such as throughput, per-flow data volume, latency, loss, and jitter, among others,that we can use to assess the network's performance on both macro- (institutional) and micro- (single researcher) scales. Thus, we can better determine the effect of changes to configurations in various components of the network. A primary metric will be the end-to-end improvements in the transport of data for specific research projects, and the ability to capture flow information and packets for analysis purposes by the campus infrastructure group and the research groups interested in networks (primarily in the computer engineering department). The monitoring and metrics put in place through this project will allow a view of network performance and key bottlenecks that is currently unavailable today, and will allow for faster, targeted remediation of network performance issues as well as proactive measures to ensure consistent optimization of data flows.

We first review the current state of the campus network infrastructure. We then present example "big science" projects that either currently or within a short period of time must manage and move large amounts of data to, from, and around the University campus. We then discuss the proposal, and emphasize that it uses existing and new equipment to design and implement a new network architecture that gives benefits over the existing one. The evaluation plans, which involve specific metrics, are in the next section. The final section reviews the goals and discusses how we will determine whether the end result meets those goals.

2. Current Campus Research Network Architecture

At UC Santa Cruz, border routers connect to Corporation for Education Network Initiatives in California (CENIC) High Performance Research (HPR) and Digital California (non-research) networks (Figure 1). These border routers then connect to core routers in a high availability configuration to a 10 Gb/s backbone matrix of connections that support very limited 1 Gb/s to 10 Gb/s research endpoint connections on campus. The architecture is highly distributed, with firewalls strategically located throughout the network such that diversified paths avoid the constraints of the security systems for trusted research that requires line rate performance, wherever they are required on campus.

Research projects addressed by this project are presently attached to the general campus production network. Research traffic competes with production campus traffic, and must be constrained to significantly less than 10 Gb/s in order not to impact campus connectivity.

The cancer genomics project has shared 10 Gb/s connectivity across the production campus backbone. The human genome browser, astrophysics cluster, and ATLAS projects are each constrained to a 1 Gb/s connection each, and must share a 10 Gb/s backbone connection with other production campus traffic.

The DYNES project, currently in progress at UCSC, will provide a single dedicated 10 Gb/s connection into the CENIC CalREN HPR-L2 network. This infrastructure can be used for scheduled dedicated 10 Gb/s connectivity, but will not support multiple simultaneous high-demand applications.

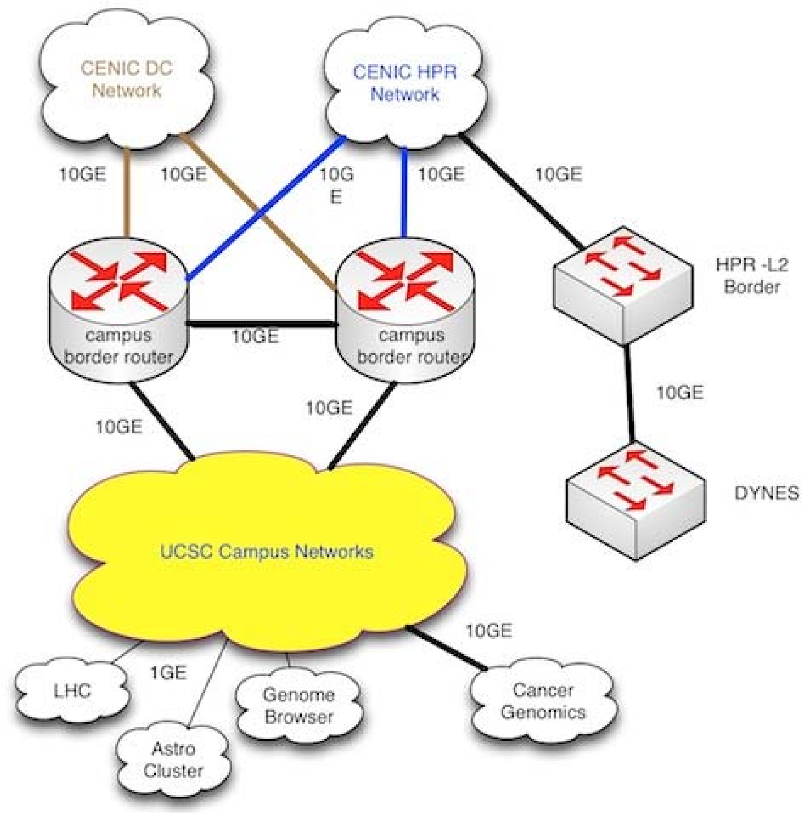

FIGURE 1. Current UC Santa Cruz Network and connected regional networks

2.1 New UCSC Campus Research Network Requirements

While the current network has historically met campus needs, significant “Big Science” research clusters have evolved at UCSC in several areas requiring a paradigm shift in their network support on campus. Currently there are researchers connecting individual clusters at 10 Gb/s rates and they forced to share them with other 10 Gb/s HPR campus users, creating a bottleneck for research.

The current UCSC network architecture is not designed to support significant new “Science DMZ” layer 2 regional, national, and international testbed networks that are isolated from the Internet, yet are massive research collaborative tools being provisioned worldwide for “Big Science” initiatives. These L2 networks include Internet2 ION, NLR Framenet, CENIC Pacific Wave and are being designed by CENIC to share the new HPR 100G physical connections being proposed with existing L3 services (see Figure 2). These new networks, in concert with the latest UCSC campus research, require the new network designs and monitoring being proposed here to avoid network data bottlenecks and delays.

The new network architecture discussed in this proposal will do much to address these needs. The enhancements to improve performance will increase throughput substantially, and thus minimize any data movement delays. As a result, existing non-research users on campus will not affect research activities. The enhancements for monitoring and network management will enable us to identify bottlenecks and compensate for them, or eliminate them entirely.

3. Participating “Big Science” Projects at UC Santa Cruz

The faculty, staff, and students at UCSC conduct research in a variety of disciplines such as the physical and biological sciences, engineering, humanities, social sciences, and the arts. Common to many of these projects is the need for high performance computing and the analysis of very large quantities of data. Often the needed computing facilities are available remotely, and the data is stored in the cloud (that is, on remote servers). As researchers collaborate with colleagues at other universities, these needs become more acute. The remote resources are now available, but they must be accessible in a way that meets the needs of the projects.

Meeting these needs is perhaps the greatest research computing challenge for the campus Information Technology Services Division.

As shown in Figure 1, UCSC is connected to the CENIC High Performance Research Network (CENIC HPR), a very high speed regional network, and thus to Internet2. A fundamental shortcoming of the current network design is that it does not provide adequate access to the CENIC HPR; that is, many researchers cannot obtain the throughput necessary for their projects over the network segment from their offices or laboratories to CENIC HPR. Worse, while some data can be stored locally, because of the mixed-use, shared backbone, collaborators will have similar problems accessing the local data -- and space at the campus data center is limited.

As examples, we discuss four areas of research that will benefit from the proposed new architecture. They all deal with the movement of large data. We emphasize that these are examples; many other projects could have been selected, because the effect of the changes will impact many research projects.

3.1 Astrophysics at UCSC

Joel Primack & Mark Krumholz

There is a large group of leading computational astrophysicists at UCSC who comprise Theoretical Astrophysics at Santa Cruz (TASC). TASC is one of the largest and most productive groups of computational astrophysicists at any U.S. university. TASC faculty include Gary Glatzmaier and Francis Nimmo (Department of Earth and Planetary Sciences) and Nic Brummell and Pascale Garaud (Applied Math and Statistics), who work on planet formation and solar and planetary astrophysics; Jonathan Fortney, Mark Krumholz, Greg Laughlin, Douglas Lin, and Adriane Steinacker (Astronomy), formation of stars and planetary systems; Enrico Ramirez-Ruis and Stan Woosley (Astronomy), stellar explosions and compact objects; and Charlie Conroy and Piero Madau (Astronomy) and Anthony Aguirre, Tesla Jeltema, Joel Primack, and Stefano Profumo (Physics), cosmology, galaxy formation, and high-energy astrophysics. UCSC is the headquarters of University of California Observatories (UCO), and many UCSC observational astronomers, including Sandra Faber and Jason Prochaska, are involved in computational astrophysics. In addition, several Physics faculty, including Robert Johnson, Steve Ritz, and David Williams are involved in gamma ray astronomy with the Fermi spacecraft or the VERITAS array of atmospheric Cherenkov telescopes. There are more than 20 UCSC postdocs in related areas. UCSC also has the new “Hyades” mini supercomputer for astrophysics, which has 3000+ cores and 12+ TB of ram. Hyades also includes the latest GPU and MIC accelerated compute nodes and an analysis/visualization node with 512 GB of ram and 32 compute cores. Hyades is used by both theorists and observers.

All of the people mentioned use powerful computers at UCSC and at national laboratories to simulate astrophysical processes and/or compare the results with astronomical observations. Many of them generate large amounts of output data. The 10 Gb/s connections that will be made possible by this grant will benefit all of this research by facilitating analysis and visualization of this data at UCSC. We will be able use the OptiPuter 40-screen visualization wall, the new 3D Visualization Lab for Astrophysics, and other UCSC visualization facilities much more efficiently with these high speed connections, and we will also be better able to prepare visualizations for education and public outreach.

Star formation provides just one example of how the 10 Gb/s connections will help. Star-forming clouds are supersonically turbulent, strongly magnetized, and radiative, and because of this complexity detailed studies of their behavior must be based largely on simulations. Because these simulations must include a wide range of physical processes over a huge range of scales, from the sizes of galaxies to the sizes of individual stars, the simulations require supercomputers and generate large amounts of data. With current bandwidth limits, moving these data sets from supercomputers to local visualization and analysis resources is not feasible, and this makes interactive study of the full data sets impossible. For example, one adaptive mesh refinement radiation hydrodynamics simulation of massive star formation [1] generated roughly 20 TB of output at the San Diego Supercomputer Center, far more than could be moved to Santa Cruz. As a result, no real-time examination of the full data set could be conducted, and analysis and visualization instead required examining isolated data frames small enough to be transmitted to Santa Cruz and then running non-interactive scripts remotely on the supercomputer. Designing and debugging the necessary code required weeks of programming. With the 10 Gb/s dedicated connection, however, a significant fraction of the data could be moved to UCSC in one day, allowing for visualization on the on campus OptiPuter and analysis using the local cluster Hyades. This would not only eliminate the need to invest resources in designing remote visualization and analysis software, it would facilitate mining of the rich data sets these simulations generate, facilitating scientific discovery.

Galaxy formation provides another example of how 10 Gb/s connections will help. The most successful high resolution galaxy formation simulation yet run was the Eris simulation [2], the PhD dissertation research of Javiera Guedes supervised by Piero Madau. Eris required more than 2M cpu-hours on NASA's most powerful supercomputer, Pleiades. The most accurate simulations of the large-scale structure of the universe yet run are the Bolshoi simulations [3] (Joel Primack, PI), which required more than 6M cpu-hours on NASA's Pleiades supercomputer, and generated more than 80 TB of output. It has only been possible to analyze a small fraction of this output at UCSC.

The University of California systemwide High-Performance AstroComputing Center (UC-HiPACC) is headquartered at UCSC, directed by Joel Primack. It links together the UC campuses and national laboratories – home to what is easily the largest and most powerful computational astrophysics faculty in the world – at a key moment when astro simulations are just now reaching their full power to explain and interpret the universe. UC-HiPACC annually organizes major international conferences and an international summer school, and facilitates intercampus and lab collaboration and outreach. All of this will benefit from the new high-speed computer connections to be funded by this grant.

3.2 Earth and Planetary Sciences
Gary Glatzmaier

Simulation studies of convection and magnetic field generation in planets and stars (G. Glatzmaier). A 3D global magnetohydrodynamic code runs on parallel computers at various NASA, NSF and DOE centers, in addition to an 3000-core cluster at UCSC. Snapshots of the simulations involve so much data that the majority of the data is stored at the centers where it is generated and only a small selected amount is transferred back to UCSC for post-processing. A 10Gb/s line would make it feasible to analyze the results more thoroughly, and would allow detailed visualization of the time-dependent 3D results.

Real-time earthquake forecasting (E. Brodsky). This planned analysis of real-time seismic observations of ground shaking will require large quantities of seismic data to be transferred continuously to UCSC from centers at UC San Diego and the Univ of Washington. A 10Gb/s line is needed for the analysis to be done quickly enough for the forecasting to be useful.

A 3D seismic reflection experiment (E. Silver). This 5-year project has just been funded to seismically image the seismogenic zone offshore of Costa Rica. The measured data will be transferred to supercomputing centers in Texas for processing and then to UCSC for analysis. Having access to 10 Gb/s connections would be of enormous benefit for our 3D seismic program. Our processed dataset is about 30 GB and the prestack data are close to 2 TB in size. Being able to transfer data at various stages of processing and interpretation with co-workers in Texas would greatly speed up our joint interpretations.

3.3 Genomics

Ann Pace

UCSC is a primary purveyor of genomics information worldwide through the UCSC Genome Browser (http://genome.ucsc.edu/), the UCSC Cancer Genomics Browser (https://genome-cancer.ucsc.edu/), and related tools and databases. These are accessed approximately one million times a week to retrieve genomic information. As part of our services, we provide access to the raw genomic data used by the browsers to researchers all over the world. This data, currently around 50 terabytes and expected to double in the next few years, is a very popular resource, making us regularly one of the top network bandwidth users in UCSC. Network upgrades to the outside world and within the campus would significantly advance UCSC's effectiveness as a central genomics resource, as we would also get quicker access to public genomic data around the world, data that we would previously not have access to due to its immense size and lengthy download time. The broader bandwidth would also greatly facilitate collaborations with the external genomics and biomedical research communities, which is awash in huge data sets generated by new experimental technologies and platforms, and struggling to cope with them.

Biomedical Genome Research

The availability of next generation sequencing data has dramatically increased the size of genomic files. These files can be up to several terabytes per sequencing run. Our research depends on the timely transfer of this information from the source (internal sequencers or external collaborators) to the places where we can store and analyze the data (compute clusters and departmental file servers). Currently, due to our limited bandwidth it can take many days to perform these data transfers. The faster the data can be moved from the sequencers to a place where they can be analyzed, such as the CBSE storage silos located in several rooms and buildings on campus, the more quickly projects will progress, thus greatly leveraging the research investment and increasing the output of sequencing facilities. This has the potential to speed up research by a factor of 3- to 4-fold. Similarly, an upgraded network to the outside world will greatly facilitate UCSC’s contribution to biomedical research through our partnerships with major medical centers such as UCSF, in which clinical sequence and other data must be transferred efficiently and securely between locations.

Also, it will be important in the near future to have high speed access to the internet from our data center in Baskin Engineering, Room 213. Currently we have a 1Gb/s link to the School of Engineering's core routers, but we would greatly benefit from a 10Gb/s link to take advantage of the extra bandwidth those routers provide to the outside world. Our cancer research relies on the ability to quickly download selected datasets from external repositories to our computing facilities to be analyzed using our compute clusters. The gathering of this data can sometimes take more time than the actual experiments. Improvements in network bandwidth would significantly reduce the time needed to collect the data, which would accelerate the pace of genomics research.

3.4 SCIPP

Jason Nielsen

Several new and ongoing research efforts at the Santa Cruz Institute for Particle Physics (SCIPP) use high-bandwidth network connections to the UCSC campus data center in addition to the external network connection to CalREN.

The ATLAS experiment at the European Laboratory for Particle Physics (CERN) searches for new elementary particles using the world's highest energy and highest intensity proton-proton collider. The experiment is expected to produce 800 MB/s of raw physics data during the upcoming run in 2015-2018. The SCIPP ATLAS group has a local computing cluster with 70 TB of reduced datasets, and the typical transfer from the nearest computing center at SLAC consists of several TB of data. These transfers currently take 1-2 days with the current network configuration, but the goal is to increase the transfer rate to 500 MB/s so that the datasets can be analyzed in a single day. The ATLAS experiment has commissioned a distributed data management system that allows local clusters to access and analyze remote datasets directly, and this access is currently limited by the UCSC network connections to the campus data center. Expanding the network capacity by a factor of 10 would bring a corresponding increase in processing rate on the UCSC ATLAS research cluster.

The VERITAS experiment is a ground-based very-high-energy gamma ray Cerenkov telescope array. The VERITAS group at SCIPP hosts some of the experiment's data, after downloading datasets from the telescope at in southern Arizona using multi-threaded network transfers. File transfers between the hosting computers and local servers often saturate the local network between buildings on campus, so the group can no longer use this high-speed transfer mode without a network upgrade.

The Dark Energy Survey (DES) and Large Synoptic Space Telescope (LSST) are two astronomical surveys that record highly-detailed full-sky maps, beginning in 2012 for DES. During the analysis phase, SCIPP researchers will analyze images in 6 GB reduced data frames from the survey telescopes. Assuming LSST analysis on a human timescale of one image per minute, the total sustained data rate between UCSC and SLAC may be as high as 1 Gbps.

3.5 End-to-End Data Transfers

Hyades users - Prof. Stan Woosley's Group. Prof. Woosley's research group in the Astrophysics Department does simulations on Type Ia Supernovae on various national supercomputers: Titan at ORNL, Hopper at NERSC and Blue Waters at NCSA. For the high-resolution simulations that they do, each checkpoint is about 280 GB and the total output per run is about 30 TB. Ideally they would like to transfer that data to their local Hyades cluster (described above) for archival and data analysis. Assuming each run takes a week to finish (which depends on how busy the queues in these clusters are), they will transfer 30TB data every week. This is for only one of the many groups using the Hyades cluster.

Cancer Genome Hub (CGHub). Network connectivity is a critical issue for CGHub. With data production projected to grow from 10 TB per month at the start of the project to 223 TB per month at the peak, the network used to connect CGHub to the participants in the TCGA project must nominally support in excess of 17 Gb/s of continuous transfer by the end of the project. This serves as a first order estimate of the overall load for CGHub, with TARGET and CGCI projects adding a modest load at a given time point. Given the expected bursty nature of the traffic, and the overhead and inefficiencies of network protocols for the high bandwidth-delay product transfers required for TCGA, this load results in a requirement for 20 Gb/s of capacity between CGHub and the Internet. While this full load must only be handled by CGHub at SDSC, the CGHub group at UCSC must be able to handle transfers of some of this load for various operational purposes as well for their role as a Genome Data Analysis Center.

Genome browser. The UCSC Genome Browser is a web-based tool that allows researchers to view all 23 chromosomes of the human genome at any scale from a full chromosome down to an individual nucleotide. The browser integrates the work of scientists in laboratories worldwide in an interactive, graphical display. Over the past year genome browser traffic has started to saturate it's current 1 Gb/s link. This project would provide the infrastructure to upgrade the genome browser's connectivity to 10 Gb/s.

3.6 ITS Collaborations

Approximately ten years ago UCSC embarked on the complete centralization of IT services across campus. One significant benefit of this consolidation was the ability of ITS to engage more effectively with researchers in providing the IT services needed to support research efforts. As a result of the consolidation ITS has had a number of notable successes in academic collaborations since its creation.

Host Genome Browser in Campus Data Center - The globally recognized Genome Browser (http://genome.ucsc.edu), run by the highly successful Center for Biomolecular Science and Engineering (CBSE), was moved to the campus data center. Prior to this move CBSE had provisioned and been managing a small data center in a School of Engineering building. This effort had been a constant source of challenges to this critical and highly visible service for the global genomics community. With the move to the campus data center these problems ceased.

Collaborate in Astrophysics Cluster - ITS partnered with the Vice Chancellor for Research, the Dean of Physical and Biological Sciences (PBSci), and faculty in the Astrophysics, Astronomy, and Earth and Planetary Sciences Departments to collaborate in the acquisition and operation of a local cluster for use in instruction and code development by these highly successful research groups. This collaboration has resulted in two successful NSF MRI proposals over the past 6 years, providing two generations of large local clusters that have significantly facilitated the success of these groups. The researchers would not have pursued these funding opportunities without ITS participation for facilities and systems support.

"CampusRocks" Cluster - ITS deployed the "CampusRocks" cluster based on donated equipment, developed and managed by the ITS staff in the School of Engineering, for use by researchers from across campus. Based on the widespread use of the cluster, the PBSci Division and School of Engineering recently committed funding for one round of upgrades to the system.

DYNES - ITS coordinated development of the successful proposal for the campus to be a participant in the DYNES project. The dynamic network services provided by this project will be an important resource for many of the network-intensive research projects on campus.

3.6. Summary

UCSC is home to a number of projects from disciplines that routinely deal with massive amounts of data. The Hyades community of users generates large amounts of data as a part of their simulation runs at various national lab clusters, and would like to be able to transfer this data back to their local Hyades cluster for archival and data analysis. SCIPP researchers depend on large data transfers as a part of their research in the ATLAS, VERITAS, DES, and LSST projects. The Cancer Genome Hub collaborates with researchers at other universities participating in The Cancer Genome Atlas project from across the United States; it receives data from TCGA Gene Sequencing Sites and sends it to TCGA Genome Data Analysis Centers on demand. Traffic to the Genome Browser, used by researchers worldwide to visualize genomes, is reaching the limits of its current 1 Gb/s connectivity.

Thus, enhancing the research network with a new architecture and more powerful links and equipment would improve the ability of researchers on campus to carry out their projects. We next discuss how we plan to accomplish this.

4. Proposed Campus Network Architecture and Milestones

This project will create a state-of-the-art networking environment for research (Figure 2) that focuses on “big data” and “big science” on campus that collaborate regionally, nationally, and internationally with world-class research initiatives. The project will implement a 100 Gb/s WAN connection, will deploy one or more 10 Gb/s connections to each of the identified research projects, and will position the campus so that successor projects can implement research network connections of up to 100 Gb/s as needs grow.

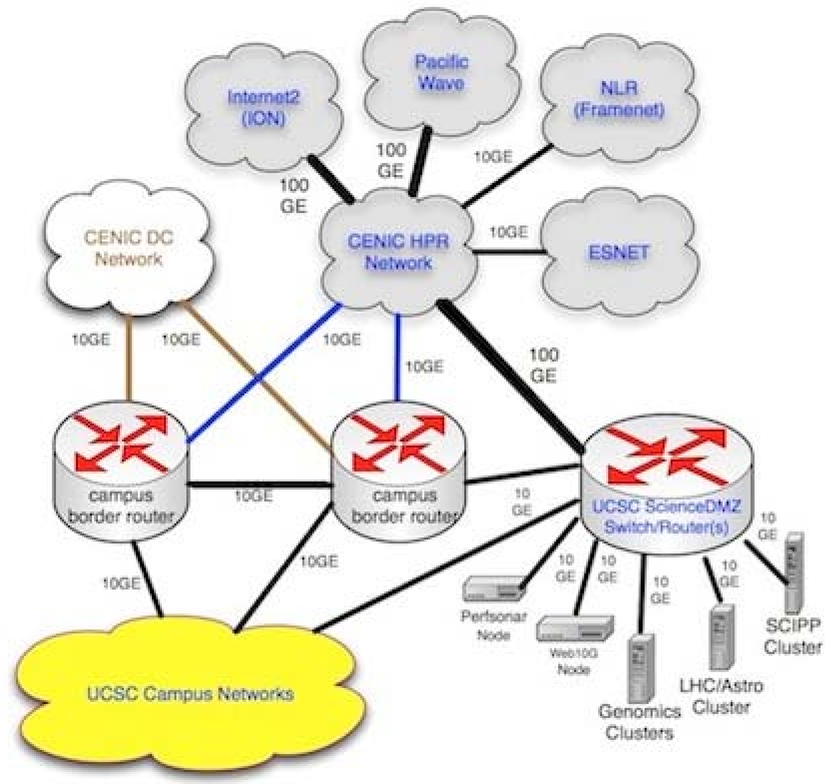

FIGURE 2. Proposed UC Santa Cruz Network and connected regional and research networks.

A new “Science DMZ” will be built on campus to interface with current regional, national, and international testbed networks being developed worldwide leveraging new L2 infrastructure from I2 (ION), NLR (FrameNET), ESNET, and Pacific Wave, to name a few. The architecture will segregate the “big science” data flows (CENIC HPR L3 and L2 networks) from more routine campus research and production network capacities. A parallel but unrelated project will refresh and upgrade the campus production network, improving connectivity for science projects of all sizes.

The redesign is focused on DWDM multiplexing at the border of campus, creating a new “Science DMZ” with high speed distribution initially to strategic research clusters on campus. The Science DMZ, a state-of-the-art router, will connect to CENIC HPR (and therefore to the world) at 100 Gb/s speeds, and distribute 10 Gb/s to campus researchers. This is effectively increasing high speed research flows through campus to the world at over 10 times the current rate. Note that this design and its growth is preserved for many years as a separate, extensive cable plant project to distribute single-mode fiber across campus is almost complete (6 months). This extended SMF fiber distribution will be leveraged by the Science DMZ in its proposed location. No additional re-engineering will be required to expand the 10 Gb/s capabilities throughout campus or to implement additional 40 Gb/s or 100 Gb/s connections as research clusters grow. The focused design provides significant value, sustainability, and growth for researchers without draining, disrupting, or distracting other network resources on campus. On campus we will verify that network performance goals are met for these significant research projects with this new infrastructure, which will include state-of-the-art network performance monitoring and troubleshooting equipment, and the latest router and switching technologies.

Off campus, UCSC will work with CENIC leadership on its Technical Advisory Council, Internet2, NLR, ESNET, Pacific Wave, and other networking partners to create and maintain effective regional, national, and international research network architectures.

Below we show specific milestones in the development of the infrastructure, including potential start dates (articulated in months, meaning months after the project begins).

4.1 State-of-the-Art ScienceDMZ [7 months]

UCSC will integrate a ScienceDMZ with this project. Currently UCSC does not have a ScienceDMZ. The ScienceDMZ is a border router/switch configuration that can support both L2 and L3 network provisioning, security, and controls. It will function to process and distribute 100 Gb/s networking technologies and multiple 10 Gb/s access and distribution segments. It will have OpenFlow support and capabilities to integrate with regional/national/international OpenFlow initiatives.

4.2 Connection To CENIC [7 months]

The UCSC 100 Gb/s connection to CENIC will leverage the CENIC HPR network architecture and partnerships to Pacific Wave, ESNET, I2 (ION), NLR (Framenet), in L2 and L3 control planes. The Science DMZ and associated router/switch hardware will control a variety of L2 control planes (depending on the L2 networks involved) and L3 HPR control plane. In addition to 10 times increased throughput, this will give UCSC researchers significant flexibility and control in creating the appropriate collaboration networks required for big data projects worldwide.

4.3 Network Test and Measurement [10 months]

The wide-area networking community views perfSONAR as an end-to-end testing facility. In this context, this means campus-border to campus border. PerfSONAR is a widely adopted tool for measuring campus to campus testing and benchmarking performance for researchers that have applications that demand high performance data rates.

Web10G fills the role of testing from within the campus to the same campus border endpoint that is serviced by PerfSonar, and can be used to test to the distant end site. It is initiated by the end user and is well suited for research entities that need to determine the state of the network before significant network applications are invoked. These tools also help a researcher determine whether performance issues are network related, computer related, or application related. Web10G also provides diagnostic information regarding protocol information on the attached system executing the test.

4.4 Establishment of an OpenFlow Network Overlay [12 months]

OpenFlow is a research platform for investigating new routing protocols, and is most promising in routing applications over the wide area. Internet2 is planning an OpenFlow testbed environment as part of its Innovation Platform. CENIC is building its California OpenFlow Testbed Network (COTN) which will interface with the I2 facility. This serves several purposes. The first is a platform for researchers to experiment nationwide with OpenFlow protocols. The second is that campus engineers can apply OpenFlow to networks to benefit from discoveries made. Finally, the campus can join nationwide test initiatives (such as GENI) that use this protocol control framework.

4.5 Support For Researchers Achieving Optimum Performance [24 months]
Research programs such as the first three described above spend much time, and incur long delays, in attempting to achieve optimum performance in transferring large data sets. Communication between the researchers and the campus central IT group is key to ensuring the enhancements to the network are done in such a way as to meet the needs of the researchers, especially when the researchers identify problems or barriers to their productivity. One of the positions this proposal will fund is a liaison whose job is to consult with the technical staff of the various research programs using the research network to select the appropriate protocols and tune systems to optimize productivity. This position would also feed information and knowledge back to the administrators of the research network, so they can make any needed changes. High performance data transfer requires specialized knowledge and experience. Providing this support centrally would ensure the liaison and administrators would benefit from exposure to many different issues, thus improve their ability to work with new or unforeseen issues. It also would avoid each research program rediscovering methods and technologies that other programs have found.

5. Broader Impacts and Intellectual Merit

5.1. Broader Impacts
The broader impacts of this work fall into two categories: impacts for research projects at UCSC, and impacts for education.

The creation of a Science DMZ with 100 Gb/s connectivity CENIC will provide a significant upgrade in the portal available to UCSC researchers for communicating with the world. This will immediately impact the existing network-intensive research being done on campus, and will provide a new resource for future research activities. This will enhance the ability of researchers here to work with external collaborators, because the campus bottlenecks arising from the discrepancy between the bandwidth of the campus network and the national networks connected to the campus (such as CENIC HPR) will be eliminated. Thus, research projects at other institutions that interact with UCSC researchers will also be enhanced.

This program will be tied to the educational programs of the associated researchers. Faculty will be able to integrate research visualizations into their classes, as well as providing access to computational systems and algorithms online. For students in most computer labs on campus or with wired laptops on campus or in residence halls, visualizations consistent with research laboratory quality would become possible. In addition, classes will be able to use the increased capacity and flexibility of the network for access to cloud-based resources. So, even though aimed at research, the proposed architecture would enhance the educational experience as well.

In addition, UCSC will provide support for an undergraduate student to be a liaison between research project staff dealing with issues of large-scale data movement and campus staff managing the research network. This student will receive training on the planned research network for science and the associated monitoring tools, and will be strongly encouraged to present their work as a poster or talk at the UCSC annual undergraduate research conference, which is modeled after professional research conferences. These activities will not only enhance their education by having them apply what they learn in class, but also will build a pool of high performance networking experts familiar with managing the motion of large data sets. As the importance of managing such data sets becomes more critical, their work will position them to continue to do research, or to work, in this critical area.

5.2. Intellectual Merit
The intellectual merit of this work lies in the planning of the new network infrastructure and architecture. How

does one make the network as flexible as possible without adversely affecting performance? How do the specific needs of a research project affect the network architecture and infrastructure? This requires an understanding of the research projects' needs, and the management structure proposed here provides a feedback mechanism to learn this.

Further, the instrumentation supports research into data management on the network, and indeed on the use of the network itself. For example, how do we optimize performance in a network carrying large amounts of data for many projects? How do we distribute our instrumentation to detect performance bottlenecks as quickly as possible? How do we work with research projects to provide them with the performance they need? How do we detect anomalies that indicate attacks, and detect attacks directly—and what is the impact on performance of such monitoring?

Finally, the question of how to bring together diverse constituencies in widely varied fields to ensure needs are met as much as possible is itself a problem of no small intellectual merit.

6. Conclusion

The University of California Santa Cruz has historically run traffic for research projects over its production networks, and has adopted a highly distributed architecture, with firewalls strategically located throughout the network such that diversified paths avoid the constraints of the security systems for trusted research that requires line rate performance, wherever they are required on campus.

UCSC is home to a number of leading research centers in science and engineering involved in network-intensive research activities. Supporting these activities at UCSC is a campus network that provides many of the features of a science DMZ. Research traffic is routed to avoid the firewalls and intrusion prevention systems that protect the administrative systems of the University. Thus, much less filtering occurs, enabling better performance and more flexibility than in a network where the traffic is routed through the administrative firewalls and other security appliances.

Nevertheless, over the past few years it has become clear that the “big data” produced and used by campus research projects will require a faster, more robust, and more flexible network on campus that is accessible by more researchers. Such a network will require new monitoring for performance and security purposes, enhancements to a number of buildings on campus where network bottlenecks exist today, and the ability to change routing quickly and easily as the traffic characteristics change.

This proposal focuses on specific technology equipment as well as specific analytical tools, that will enable our researchers to more effectively exploit the rich range of capabilities available from high performance networks, and collaborate with colleagues throughout the scientific world. To realize this vision, we also propose a strong collaboration of researchers with the central campus IT group, investments in an expert in performance analysis, and support for student engagement with the researchers. Because of our experience with the distinctive partnership of researchers and the central campus IT group, the investments outlined here will have substantial value for the research community as we work to address the specific needs of our existing research programs and develop a sound strategy for improving the campus cyberinfrastructure to address even larger-scale data-intensive applications over the next 3 to 5 years. Investments proposed here will immediately impact our high-performance research networking infrastructure. Finally, we have articulated a plan that, while focused on demonstrable and measurable value on campus, is positioned to create similar support for multi-institutional research initiatives, notably through our robust partnerships with major networking partners, as well as sister research universities.

[1] Krumholz, M. R., Klein, R. I., McKee, C. F., Offner, S. S. R., & Cunningham, A. J. "The Formation of Massive Star Systems by Accretion", 2009, Science, 323, 754.

[2] Guedes, J., Callegari, S., Madau, P., & Mayer, L. " Forming Realistic Late-type Spirals in a ΛCDM Universe: The Eris Simulation", 2011, Astrophys. J., 742, 76.

[3] Klypin, A. A., Trujillo-Gomez, S,, & Primack, J. "Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation", 2011, Astrophys. J., 740, 102. Links to other journal and popular articles about Bolshoi and visualizations are at http://hipacc.ucsc.edu/Bolshoi.

Budget Justification

A. We are requesting 1 month of support for each year for Dr. Brad Smith (Director Research and Faculty Partnerships, ITS), who will provide advisory oversight to the project during the academic year with specific effort during the summer to oversee the designing and tuning of the network equipment, monitoring tools, and support process.

B. We are requesting funding for 1 undergraduate student to receive training on the planned research network for science and the associated monitoring tools in order to provide support and expertise to campus researchers. The students will work under the direction of the HPC Network Engineer to support the planned science research network including advising and assisting campus researchers in the use of the scienceDMZ research network and any performance tuning needed as identified by network monitoring tools. Undergraduate researchers may be able to conduct this work partly as upper division coursework and will be strongly encouraged to present their work as a poster or talk at the UCSC annual undergraduate research conference, which is modeled after professional research conferences.

We are also requesting support for a highly experienced High Performance Computing (HPC) Network Engineer, who will be a critical component to this project is this person will architect the proposed research network, implement and tune the monitoring tools, and establish and train undergraduate students to support campus researchers. The HPC Network Engineer must have experience designing leading edge research and traditional networks, understand the needs and challenges of campus researchers. The HPC Network Engineer will serve as the project lead in designing and implementing the proposed research network, network monitoring tools, and support processes. Our plan is to use this funding to provide network performance support to the growing number of network-intensive research projects.

D. The equipment identified here will be used to build the enhanced research network. See the Project Description for details about how the equipment will be used.

Facilities, Equipment and Other Resources

The UC Santa Cruz campus covers 2000 acres overlooking scenic Monterey Bay. The campus is served by an aging fiber optic and copper cable plant that will be substantially replaced by the Telecommunications Infrastructure Upgrade (TIU) project over the next five years. Campus backbone data connectivity is provided by two core routers that are connected at 10 Gb/s to eleven distribution routers at Area Distribution Facilities (ADFs) in different campus geographic areas. The TIU project will upgrade ADF facilities, outside plant, building terminal rooms (TRs) and horizontal cabling to support converged delivery of data, voice and video services over IP to end stations at a baseline rate of 1 Gb/s with power over Ethernet available anywhere. Outside plant upgrades will augment single mode fiber into each included building.

The two campus core and minimum point of entry (MPOE) facilities house campus border and core routing equipment that is configured such that campus connectivity will not be harmed by the failure of a single core/MPOE. These two facilities are now about 300 meters apart, in the Communications (Comm) and Interdisciplinary Science (ISB) buildings. The TIU project will relocate the ISB core facility to a new MPOE building at the base of campus, providing physical separation of about 3 km between the two core/MPOE facilities. TIU outside plant work will provide each ADF with an independent fiber path to both core facilities. The secure core/MPOE facilities will be provided with generator and battery-backed power to survive utility power outages without impact to active electronics or cooling equipment.

The campus has two California Research and Education Network (CalREN) points of presence provided by CENIC. The CalREN network utilizes Dense Wave Division Multiplexing (DWDM) over a fiber optic path from UCSC to Sunnyvale to deliver network connectivity 10 Gb/s, and can support multiple future 100 Gb/s connections, one of which would be dedicated to the use proposed here. This wide area connectivity is terminated in equipment located at the Physical Sciences (PSB) building, from where multiple 10 Gb/s connections are extended to both core facilities. Backup connectivity from UCSC to CalREN in Oakland is provided by a 10 Gb/s AT&T Decaman circuit that is provisioned over a physically diverse path.

The campus network is currently connected to remote locations and the Internet in general by CENIC's CalREN Digital California network, which also connects virtually to all of the K-12 school districts, community colleges, the twenty-three California State University campuses, ten UC campuses and five medical centers, and three private research institutions within the state. This connectivity is provided by a 10 Gb/s channel. There is also a 10 Gb/s link into the CalREN High Performance and Research (HPR) network, and a 10 Gb/s link into the CalREN HPR-L2 experimental layer 2 network.

In the existing campus production network firewalls are implemented at the edges to protect sensitive data. There are many current methods of campus research network connectivity; all of them involve use of departmental or building networks that share the campus 10 Gb/s backbone and the 10 Gb/s WAN links into CalREN. The experimental DYNES project is the single exception, having one 10 Gb/s connection to CalREN, thence into Internet2 ION.

The proposed project would take advantage of a CENIC-funded 100 Gb/s interface into the CalREN HPR network by provisioning a switch that would be able to provide dedicated 10 Gb/s bandwidth to multiple UCSC research projects, while protecting both science and production campus traffic from mutual contention on the campus backbone and CalREN WAN connections.

Data Management Plan

Data gathered in this project will consist of performance measurements and other metrics about the use of the network (including, potentially, network traces and captured packets). In some cases, these data will contain information that the campus deems personally identifiable.

Accessing and Sharing Data

In general, only central campus IT personnel will have access to the raw data. This preserves the privacy of the information gathered. In some cases, researchers may seek access to the data. A protocol similar to the following will be used.

All data will be held under the control of the central campus IT group. If the data the researchers want to access is deemed non-sensitive, then the researchers will be allowed to access the data subject to the rules of the central campus IT group. If the data is deemed sensitive, the researchers must first obtain approval (or exemption) through the Institutional Review Board. They may then access the data on the repository according to the protocol approved by the IRB and central campus IT group. They may not remove the sensitive data, but may analyze the data and remove the results (assuming the IRB has approved this). This preserves privacy, while allowing access to researchers who can appropriately justify their need.

Both the IRB and the central campus IT group must approve any publication or sharing of the raw data with the general community.

The performance measurement data will be captured in a form that the specific tools can use. With packet capture and network flow data, we will use a standard form such as pcap and netflow or flow tool format. This will enable us to use standard tools to analyze the data.

Preserving Data

The captured data will be kept as long as possible on the central campus IT collection system. When more room is needed, the data will be overwritten.

UCSC Research Computing Infrastructure Plan

Introduction

In recent years, increasing recognition has been devoted to the need for U.S. research universities to develop a balance of coordinated and centralized information technology (IT) capabilities for computationally based, highly collaborative research. In particular, a 2003 report from a National Science Foundation (NSF) blue ribbon panel chaired by Prof. Dan Atkins of the University of Michigan (and later the first director of the NSF Office of Cyberinfrastructure) presented a broad plan for making strategic investments in CI technologies and services across the research disciplines and coined the term cyberinfrastructure:

> *Like the physical infrastructure of roads, bridges, power grids, telephone lines, and water systems that support modern society, "cyberinfrastructure" refers to the distributed computer, information and communication technologies combined with the personnel and integrating components that provide a long‒term platform to empower the modern scientific research endeavor.*

Recently, the EDUCAUSE working group on Campus Cyberinfrastructure has advanced the following definition of CI:

> *Cyberinfrastructure consists of computing systems, data storage systems, data repositories and advanced instruments, visualization environments, and people, all linked together by software and advanced networks to improve scholarly productivity and enable breakthroughs not otherwise possible.*

At the national level, CI planning and investments have been driven by the growing complexity of modern IT, the importance of distributed collaboration, and the increasing IT demands in the research disciplines within the physical and social sciences, engineering, and medicine. In addition to the aforementioned efforts at NSF, other major federal R&D funding agencies including the National Institutes of Health (NIH) and the Office of Science within the Department of Energy (DoE) have placed increasing emphasis on identifying CI needs within their supported disciplines and are making initial investments.

These efforts are all based on the understanding that scientific, engineering, scholarly, educational activities are deeply impacted by developments in information technologies. These innovations improve upon established approaches as well as make it feasible to follow previously unexplored paths to manipulate and investigate 'big data' in multiple ways. To ensure that its researchers, scholars and students remain productive and competitive in the coming decade, a university must provide the necessary framework and tools.

Analysis of UCSC's IT Environment

Following is a brief SWOT analysis of UCSC's existing cyberinfrastructure with the goal of identifying the actions we can take that will provide the greatest ROI in improving the CI available to UCSC researchers. In the following we identify existing strengths that ITS can build on, weaknesses we need to address, opportunities (i.e. potential future strengths) we might exploit, and threats (potential sources of future weaknesses) we might address. Based on this analysis we identify a set actions we propose for developing our IT infrastructure over the coming 5 years.

<u>Strengths</u>

- **Presence of Highly-Successful, IT-intensive Research Activities** - As with any research university, the heart and soul of the institution is its academic enterprise. A major strength of UCSC, that is a primary focus of the Information and Technology Service (ITS) Division's support efforts at UCSC, are the highly-successful, IT-intensive research activities existing on the campus. Fortunately, UCSC is home to a number of such activities.

  - There is a large group of leading computational astrophysicists at UCSC who comprise Theoretical Astrophysics at Santa Cruz (TASC). TASC is one of the largest and most productive groups of computational astrophysicists at any U.S. university.
  - The Center for Biomolecular Science and Engineering (CBSE) promotes and supports genomic and stem cell research, technology innovation, and education.
  - The Santa Cruz Institute for Particle Physics (SCIPP) is an organized research unit within the University of California system. SCIPP's scientific and technical staff are and have been involved in several cutting edge research projects for more than 25 years. The primary focus is experimental and theoretical particle physics and particle astrophysics. The Institute is recognized as a leader in the development of custom readout electronics and silicon micro-strip sensors for state-of-the-art particle detection systems.

- **Fully Centralized IT Organization** - Ten years ago UCSC embarked on the complete centralization of IT services across campus. As a result, all IT services now report to the Vice Chancellor of Information Technology Services (VCITS). This is the most aggressive consolidation of IT in an academic environment we are aware of. This consolidated organization provides agility in delivering IT services to the campus, and more efficient and effective use of IT resources in addressing user needs.
- **Strong and Successful Collaboration between ITS and Research** - Two innovations implemented as a part of the IT consolidation described above have enabled the resulting ITS organization to provide *improved* support to the research and instructional activities on campus: the Divisional Liaison (DL) and Local IT Specialist (LITS) roles in the academic Divisions; and the Research and Faculty Partnerships (RFP) unit reporting to the VCITS.

  The DL is responsible for managing the IT services provided to a Division and supervises the LITS allocated to that Division. The DL has a dual reporting relationship; 50% to the VCIT and 50% to the Divisional Dean. The DL/LITS support model for the academic Divisions has allowed the new ITS organization to realize the benefits of a consolidated IT organization without sacrificing the attention to local needs that is often feared of consolidation.

  The RFP unit is composed of two IT professionals with strong academic backgrounds, and long history with the campus. The RFP Director has an Adjunct appointment in the Computer Engineering Department, and the Senior Network Engineer in RFP has a PhD in Chemistry and was the Senior Network and Systems Engineer for the campus prior to his assignment to RFP.

  Based on these consolidated academic IT support units, ITS has had a number of notable successes in academic collaborations since its creation. Soon after consolidation, ITS identified the lack of dedicated fiber connectivity to the campus, and the critical nature of this connectivity for future research projects. Working with campus faculty and academic administration, the regional network provider (the Corporation for Education Network Initiatives in California, or CENIC), and the UC Office of the President (UCOP), ITS developed a project proposal for the fiber build, acquiring funding commitments from UCOP and the campus, and worked with CENIC to manage the implementation of a fiber path to the campus.

  ITS worked with CBSE to move the globally recognized Genome Browser (http://genome.ucsc.edu), run by the highly successful Center for Biomolecular Science and Engineering (CBSE) to the campus data center. Prior to this move CBSE had provisioned and managed a small data center in a School of Engineering building. This effort had been a constant source of challenges to this critical and highly visible service for the global genomics community. With the move to the campus data center these problems ceased.

  ITS partnered with the Vice Chancellor for Research, the Dean of Physical and Biological

Sciences (PBSci), and faculty in the Astrophysics, Astronomy, and Earth and Planetary Sciences Departments to collaborate in the acquisition and operation of a local cluster for use in instruction and code development by these highly successful research groups. This collaboration has resulted in two successful NSF MRI proposals over the past 6 years, providing two generations of large local clusters that have significantly facilitated the success of these groups. The researchers would not have pursued these funding opportunities without ITS participation for facilities and systems support.

ITS deployed the “CampusRocks” cluster in 2009, developed and managed by the DL for the School of Engineering, for uncompensated use by researchers from across campus. Based on the widespread use of the cluster, the PBSci Division and School of Engineering recently committed funding for a first round refresh to the system.

ITS served as the Network Engineering consultant to CBSE in the development of a proposal to NIH to develop the Cancer Genomics Hub (cghub.ucsc.edu). This project involves archiving cancer genomes sequenced by medical researchers around the country, and accessed by computational genomics groups from around the country. This network-intensive project would not have been possible without the network expertise from ITS and the availability of dark fiber described above.

ITS coordinated a successful proposal for the campus to participate in the DYNES project. The dynamic network services provided by this project will be an important resource for many of the network-intensive research projects on campus.

- **Proximity to Silicon Valley** - UCSC is located ~20 miles south of the world famous Silicon Valley. Silicon Valley is home to the world’s most successful high-tech ventures, and there has been a strong and developing set of collaborations between UCSC researchers and Silicon Valley companies. This is a resource ITS hopes to continue developing in the future.
- **Existing Cyberinfrastructure Strengths** - UCSC has two notable cyberinfrastructure strengths to build on: its existing datacenter service, and its external network connectivity. With limited resources and a geographically challenging environment, UCSC has had to innovate in the development and management of its data center. The existing data center has been near capacity for approximately eight years. To deal with continuing growth in data center needs by the campus, ITS has pursued two innovative solutions to address these needs with our limited resources. First, we were an early adopter of server virtualization technologies that have allowed us to support a steady growth in the campus’s server count during this time without increasing the data center’s power and cooling capacity. Second, UCSC has been an early and aggressive user of consolidated data center services around the UC system. Specifically, we have made aggressive use of capacity allocated to UCSC at the San Diego Supercomputer Center, allowing us to expand the power and cooling capacity available to campus researchers without the need to expand campus facilities.

  As a result of the dark fiber project described earlier, UCSC has robust, cost-effective, and high-performance network connectivity to regional, national, and global research networks. In addition, UCSC is an active participant in the governance of the Corporation for Education Network Initiatives in California (CENIC).
- **Existing IPv6 Deployment** - UCSC deploying IPv6 campus-wide. The campus has a /32 IPv6 allocation from ARIN, and has a detailed plan for IPv6 addressing on campus. The /32 is publicly advertised via BGP to CENIC. IPv6 is routed in the campus backbone and is currently available on subnets within the Communications building. Campus perfSONAR nodes are IPv6-enabled allowing for monitoring and measurement of the IPv6 connectivity to remote participating perfSONAR sites. DHCPv6 is used for host addressing. Netflow data is used to provide visibility into traffic flow behavior utilizing the Nfsen/Nfdump software.
- **Existing InCommon Deployment** - As a member of the InCommon Federation, UCSC provides Shibboleth authentication services based on the UCSC Cruz ID and what we call a Gold password. We currently are not InCommon Bronze or Silver certified. We are working towards an InCommon Silver certification. The planning and funding for this effort is not yet in place. UCSC also utilizes the

InCommon Certificate Service for managing SSL/TLS certificates.

Weaknesses

- **Geographically Challenged** - UCSC is geographically challenged in the sense that the northern Monterey Bay region is surrounded by coastal mountains and environmentally sensitive wetlands that provide significant obstacles to the development of critical infrastructure services such as communications and power. In a sense, the wealth in natural resources and beauty that have brought the region worldwide fame are ironically the source of one of its major challenges for the development of these critical infrastructures. This has led to challenges. The UCSC fiber build was longer, and more expensive by an order of magnitude than any other California research university. Geography has also been a source of the difficulty of increasing data center capacity on campus.
- **In-building Cable Plant** - Due to long-term under-investment, UCSC's in-building cable-plant is severely out-of-date. Many buildings are limited to 100 Mb/s, with some limited to 10 Mb/s.
- **Diminishing State Support** - With the recent economic downturn, UCSC in general, and ITS in particular is still feeling the effects of an extended period of painful budget cuts. The ITS budget has been cut approximately 25% over the past four years with the result that we are having to meet a continuing growth in demand with dramatically reduced funding.

Opportunities

- **Campus Commitment to Telecommunications Infrastructure Upgrade** - ITS has developed a comprehensive plan for upgrading the campus communication infrastructure over the next seven years. Funding has been committed to this project, and early phases of the project are underway.
- **Cloud Computing** - The availability of cloud-based computing services, including the San Diego Supercomputer Center, provides a potentially scalable and cost-effective means for continuing to provide the ITS infrastructure the campus needs during a time of diminishing resources.
- **Growth Rates in IT Capacity** - Moore's Law (transistor density in integrated circuits doubles every 18 months) and Nielsen's Law (network bandwidth increases 50% every year) provide the promise that through normal replacement cycles we will experience significant growth in IT capacity.
- **Software-Defined Networking and OpenFlow** - These technologies offer the promise of providing more efficient use of networking resources, and thereby getting an improved return on investment in networking resources.
- **Offer of 100 Gb/s upgrade from CENIC.** CENIC is offering its high performance research community 100 Gb/s upgrades to their HPR network connections.

Threats

- **Further Budget Cuts** - The State of California is still wrestling with revenue shortfalls, and the University of California continues to expect further reductions in funding levels in the coming years.
- **Lack of Physical Diversity in Dark Fiber Infrastructure** - With its challenging geography, UCSC continues to wrestle with connectivity challenges. While we have solved the basic dark-fiber challenge, we are now wrestling with the physical diversity challenge.
- **Energy Costs** - Power costs in Santa Cruz are high compared with other parts of the state and nation.
- **Data Center is near Capacity** - While robust and efficiently run, UCSC's data center is near capacity.
- **Growth Rates in IT Capacity** - Paradoxically, this opportunity is also a threat. For researchers to stay competitive in the IT-intensive fields requires staying at the leading edge of these growth rates. The extremely short generational cycles in these technologies requires a very aggressive effort to keep up with the requirements of these disciplines.
- **Competitive Job Market** - Being located close to Silicon Valley, home to the greatest concentration

of high-tech innovation and entrepreneurship in the world, UCSC is constantly challenged with hiring and retention of highly skilled IT staff in the presence of the nearby, insatiable job market.

UCSC's CI Plan

Based on this strategic analysis, ITS is engaged in the following efforts to develop UCSC's Cyber Infrastructure:

Develop Computational Capacity - Data center services review with goal of assessing different options for data center capacity, including the use of cloud services, expanding the use of regional UC data centers, and expansion of the campus's data center.

On-Campus Network Upgrade - The campus has committed to the Telecommunication Infrastructure Upgrade, which will involve upgrading campus cable-plant and network equipment, and adding inter-building fiber paths to provide the robustness needed for this critical infrastructure.

Off-Campus Network Connectivity - ITS is pursuing a number of options for building a second, physically-diverse fiber path to the campus. In addition, ITS is pursuing funding to continue upgrading campus connectivity as new technology becomes available and cost effective.

Research Support Capacity - To expand the research support capacity to meet the growing needs of existing IT-intensive disciplines, and to expand such support to other disciplines, ITS is pursuing external funding to continue existing activities and provide seed funding for additional ones. We have been successful with support for the Astrophysics cluster, which now has staff support funded from largely external sources for the next 5 years. A possible model to pursue is the *Research Liaison* successfully used with the Astrophysics cluster support.

IPv6 - Campus-wide deployment on user-facing subnets will begin in the summer, 2013. Some departmental public-facing services will deploy IPv6 in late 2013. The main campus public and internal services, most of which are located within the data center, will eventually have native IPv6 access, but await a firewall replacement project and subsequent planning and design. IPv6 access to these services might be provided via translation in the interim.

InCommon - In 2013, the campus will begin ordering SSL certificates for servers via the InCommon service. Following that work, UCSC will be evaluating the use of the InCommon Certificate Service to manage Client/Personal Certificates and Code Signing Certificates.