UNC-Chapel Hill CyberInfrastructure Plan Summary

The University of North Carolina at Chapel Hill (UNC-CH) recognizes three categories of strategic effort associated with cyberinfrastructure:

1. Foundational capabilities
2. Collaboration enablers
3. Purpose-specific enhancements

UNC-CH views (1) as being predominantly institutionally focused and (3) as predominantly project or discipline focused. While (2) assumes basic nationally shared functions and technology standards, it also includes services that are established by disciplinary communities of practice.

---

1. *Foundational capabilities*

There are three capabilities that serve as foundation: (i) high capacity and high-IO data storage, (ii) high-bandwidth and low latency networks, (iii) computational power. These are core capabilities for current scientific and health research; and they serve as springboards for next generation projects. It is imperative that UNC-CH focus investments in these areas.

Initiatives:

a. Refactor LAN security design to remove bottlenecks
b. Upgrade deficient access-layer network endpoints to 1Gbps minimum standard
c. Upgrade aggregated network links to eliminate transmission throttling
d. Plan for the implementation of IPv6
e. Upgrade shared services high-end computational capability
f. Expand storage services and file system offerings to meet varying capacity, performance and throughput requirements

The University of North Carolina at Chapel Hill (UNC-CH) implements a network infrastructure strategy that balances between four competing desiderata: (i) minimization of latency, (ii) maximization of throughput, (iii) fault containment, (iv) security. With no structurally guaranteed capital infrastructure/lifecycle funding since 2006, UNC-CH has prioritized ad hoc investments, typically spawned by emergent needs, per those desiderata. Emergent needs centered principally on security.

Prior to 2010, there was disproportionate emphasis on intra-network security controls: this desideratum put the others in shadow (with fault containment a distant second). Because the intra-network security controls could sustain line-rates of only 1Gbps, network flows to endpoints were necessarily constrained. This, in turn, stunted relevance of improving line-rates to endpoint devices. Nearly 44% of UNC-CH's endpoint ports are constrained to 100Mbps, either because that is the maximum line-rate of access layer switching gear, or because building entry switches can aggregate only up to 1Gbps. Partly because extramural investments demonstrated significant scientific value, UNC-CH has committed to an ongoing lifecycle funding model for core network infrastructure that is on a multi-year phase-in.

Since 2010, UNC-CH has refactored the network architecture based upon actual analysis of threat vectors and placed its principal security devices at the network borders. As a result, UNC-CH has been able to remove most of the intra-network security devices, and in so doing eliminate the unnecessary constraints that stunt both latency minimization and throughput maximization. A concomitant result has been that

researchers with data intensive or network intensive needs have driven access layer devices to saturation, and many upstream devices to saturation. At worst, this stymies the science such researchers pursue; at best, it slows.

UNC-CH has two /48 assigned IPv6 global unicast address networks; at the present time, the first /48 is advertised to the Regional Optical Network (RON) NCREN, while the second /48 is used for on-campus only traffic. IPv6 address spaces are allocated across campus based on location, network aggregation points (routing domains), and address space functionality. Networks up to /52 are allocated for each routing domain for easy address summarization by relevant routing protocols. Our emphasis has been directed towards IPv6 deployment via dual-stack on the most important and popular public-facing campus servers, including www.unc.edu, help.unc.edu and ibiblio.org, as well as all campus DNS servers. We have also published a page of information and testing tools for IPv6 (http://ipv6.unc.edu) that is accessible over both IPv4 and IPv6.

In collaboration with Duke University and North Carolina State University, the strategy for border links includes support for NCREN to deploy 100Gbps capable network transport to the regional optical network. Although demands for border links are presently driven by aggregate bandwidth demands rather than research demands, UNC-CH is planning its border strategy for an emergence of research-oriented volumes/flows. Accordingly, UNC-CH aims to align its intra-campus networking strategy to the eventual high-throughput border links.

UNC-CH has recently procured replacements for its aging high-end compute facility. New capabilities include a 772 node (9152 core) Dell Linux cluster, “Kill Devil,” with QDR Infiniband interconnect and a minimum of 4 GB memory per core; a smaller 2300-core HP Linux cluster with QDR Infiniband interconnect and at least 6GB of memory per core; and two 32-core hosts with one terabyte of memory each to accommodate codes that require extremely large amounts of RAM. The Kill Devil Cluster also includes 64 NVidia Tesla GPUs (M2070). Storage for research data includes more than 2.5 petabytes of disk, comprising locally attached disks; network-attached shared scratch space; and network-attached shared file systems. Supported file systems include Lustre, Isilon OneFS, GPFS, and NFS. A variety of services are also provided to accommodate data transfer, processing and storage capabilities for sensitive data. In acknowledgement of the ongoing growth of data-intensive sciences, UNC-CH has procured an advanced mass storage archival system that replaces the legacy 700TB with a 2PB service – and improves performance by at least an order of magnitude. UNC-Chapel Hill has established a new funding model to sustain this high-end compute facility moving forward. The funding model is expressly designed to open high-end computing to new research endeavors: it is structured to be funded with 70% central/institutional funds, and 30% utilization fees to users. Utilization fees are assessed once researchers exceed a default allocation funded by the central/institutional funds.

2. *Collaboration Enablers*

Scientific and health research continues to become ever more multi-disciplinary and multi-institutional. As such, it is essential to continue to foster UNC-CH’s engagement with the Internet2 middleware stacks to further opportunities to establish federation relationships with other research entities, and to refine group and user management, provisioning, and deprovisioning, services.

Initiatives:

a. Participate and enhance Incommon qua Shibboleth federation
b. Pursue multi-institutional group-management
c. Pursue next generation user/account provisioning/deprovisioning solution

d. Implement multi-agency data sharing solution
e. Build appropriate research data stewardship guidelines and recommendations

UNC-Chapel Hill selected Shibboleth as the Web Single Sign-On solution and first implemented it in 2008. Since that time, there has been widespread adoption of Web SSO across campus. Today there are nearly 200 Service Providers interacting in a non-federated fashion directly with the sso.unc.edu Identity Providers, and many more federating with UNC-Chapel Hill through the InCommon, UNC-General Administration or NCTrust federations. Some prominent applications that use Web SSO are ConnectCarolina Campus Solutions, Blackboard, Sakai, ePro, Library EZProxy, and UNC Virtual Computing Lab. In 2011, the Identity Management team upgraded the Shibboleth IdPs to include the ECP extension in order to allow authentication with the Microsoft Live@EDU solution for student email. The IdPs now handle nearly a million logins per day.

For authorization and groups management, the Internet2 Grouper solution has been implemented. Grouper provides both user-interface-based and automated groups management. The "dynamic" groups capability allows for adding and removing users from groups based on a change in department, affiliation, or other identity attribute. These group memberships are then populated in LDAP and Active Directory, and can also be sent with the SAML assertion through Web Single Sign-On. Grouper is becoming more widely used across campus, especially since the addition of POSIX groups management to the Grouper solution. Early adopters of Grouper have included the Library for eduPersonEntitlement membership population, and the Carolina Digital Repository, where Grouper is used to allow certain users access to specific repositories.

In the provisioning space, the Identity Management team is deploying the first phase of a system called IMPROV (Identity Management Provisioning). IMPROV consists of a Service Provisioning Markup Language (SPML) -based router mechanism that interacts with individual Services that provide our login identifiers, the Onyen and the UNC Guest ID. We intend future phases to include De-provisioning for these identifiers, and Provisioning/De-provisioning for other services such as Live@EDU and Exchange. We receive many requests across campus for a way to alert departmental IT groups of the departure of employees so that access to services can be removed. This will be our means of providing de-provisioning information across campus.

3. *Purpose-specific enhancements*

While some disciplines know well the return computational methods deliver to research, others know less well. UNC-CH therefore should follow a twofold strategy: (i) pursue partnership opportunities to share investment in specific research projects that will benefit larger communities of users, (ii) pursue outreach and engagement activities with disciplines less familiar with computational techniques.

Initiatives:

a. Provide common and industry-standard software and libraries
b. Establish staffing support to engage researchers in the outreach disciplines
c. Design and deploy a “secure research environment” for human subjects research
d. Provide “stepping-stone” heterogeneous computing capabilities
e. Facilitate access to national cyberinfrastructure resources

Centrally provided and managed software applications include a variety of Fortran and C compilers; Gaussian, Amber, Insight II, and dozens of other open-source packages commonly used in the biosciences

and physical sciences. In addition, standard mathematical and statistical software such as Mathematica, Matlab, Stata, R, and SAS are available, as are GIS applications, including ArcGIS, and visualization/image processing packages such as ENVI and IDL. Virtual computing software developed at North Carolina State University has been deployed on many nodes of the general purpose compute cluster to allow researchers to run Windows or Linux-based applications easily on central computing systems. Computational scientists on staff in Research Computing are available for consultation on use of systems, applications, modeling, analysis and code optimization.

In addition, Research Computing provides and supports a CentOS-based Linux image (TarHeel Linux), customized for use with the UNC technology infrastructure, and a software application repository to enable rapid deployment of distributed Linux workstations on campus. Codes developed and tested in these environments can easily scale up to the central compute clusters.

In an effort to broaden the reach, applicability, and fecundity, of computational methods, Research Computing at UNC-Chapel Hill has established outreach and support staffing to foster computational techniques in the services of research in the arts and humanities. This extension effort has been in place since February 2011; additional extensions are under consideration. These extensions are to augment our core engagement team that includes advanced computational scientists with deep expertise in chemistry, physics, mathematics, bioinformatics, and other areas.

SMRW, the Secure Medical Research Workspace, is a comprehensive solution to protect Electronic Health Records (EHR). The SMRW utilizes virtualization technologies to facilitate the setup, data provisioning, management, and tear down of protected virtual workspaces. The virtual workspace incorporates Data Loss Prevention (DLP) technologies and techniques to prevent unauthorized use and transmission of data in order to maintain compliance with Institutional policies and HIPAA data regulations. A generalization and extension of SMRW for high- and medium-risk human subjects research, called the “Secure Research Environment,” is in planning and pilot phases.

The Kill Devil Cluster’s 64 NVidia Tesla GPUs (M2070) are available for researchers to explore programming in CUDA or OpenCL for projects that significantly benefit from, e.g., parallel processing and floating point calculations.

Condor-based grid capabilities have been deployed centrally and in various campus units, especially RENCI; Research Computing staff facilitate access to these, other Open Science Grid (OSG) services, and XSEDE national resources. RENCI has served as the OSG Engagement Coordinator, working with researchers from various disciplines to integrate their applications and workflows into the OSG framework when campus computational resources are insufficient to meet specific needs. Research Computing has donated 560 compute cores to RENCI for OSG use, replacing less capable resources. XSEDE campus champions Mark Reed and Michael Barker also provide assistance in accessing XSEDE resources and opportunities.