Research Computing Advisory Committee

Minutes Jun 3, 2013 (taken by Erik Deumens)

Present: Erik Deumens, David Nessl, Dave Pokorney, Alberto Riva, Laurie Taylor

Discussion

Unified campus network project

After the investment in upgrades to the campus research network this winter and spring, the University of Florida is starting a project to design and implement the next generation architecture for a unified campus network, including the buildings housing researchers managed by CNS and HealthNet, that will allow optimal use of the new research network infrastructure with minimal friction.

As part of this process, comments and input from the community of faculty and researchers is sought as to what the needs are and what the points are where friction exists that should be reduced. Input was received by email during the two weeks before this meeting and further input was received during discussion at the meeting.

The discussion focused on concrete problems from having the current separated research networks with main campus and the Health Science Center:

- researcher frustration on things that don't work or are cumbersome to do,
- problems with compliance with policies that do not have associated tools and infrastructure needed to implement or comply with them,
- workarounds are found to get things done, with the resulting disengagement by the researchers from the IT providers on campus,
- confusion over network overall with the split,
- confusion overall with the network not being transparent,
- lost opportunities with even de-identified health data not easily accessible by researchers

There is a recognized need to protect systems with health data from the campus network that includes the dorms, since a large number of malware comes from that part of the campus network. However, new technology and advanced maturity of technology now allows a more fine-tuned protection to be built around systems with health data without the need to work with entire buildings as indivisible units. Virtual routing and forwarding (VRF) is now available on all major router equipment and can be used to build much more configurable environments without compromising on security.

Data sharing for collaborative research is essential for UF to become a top 10 university. Ideally users should be able to specify data flows between logical entities for acquiring, storing, analyzing, archiving data and provide annotations in metadata with

details. Then the infrastructure will make sure that data moves and is processed safely and efficiently.

With the existing infrastructure with multiple firewalls and VPNs, researchers must connect to one server with one VPN copy data to their workstation, then connect to another server with another VPN and copy the data. Now a copy of the data is left in the worst possible place: the user workstation.

The services providers supporting research activities on campus need to collaborate to create a coherent environment for researchers from their point of view, not from the point of view of the providers: network providers, storage and processing providers (labs, departmental IT, college IT, AHC IT, CNS, Research Computing, libraries).

HIPAA compliant HPC and big data infrastructure

Erik asked about HIPAA compliance regulations and any specific technical requirements, which led into a discussion on policies, procedures, and compliance with unification and Sarbanes-Oxley, FERPA, HIPAA, and other requirements that have to be met.

The basic requirement in the law for HIPAA and FIRPAA is to make sure that reasonable measures have been put in place to secure the data and to ensure that data does not get into the hands of people who have no rights to it or no need to see it. The newer HITECH law adds the requirement to provide an audit trail that show who accessed data when.

When implementing HIPAA and HITECH compliant systems, it is useful to look at a set of rules that was created as a checklist: SAS70 (which was really created around Sarbanes-Oxley) is helpful, especially type II applies to HIPAA compliance. SSAE16 is a new version.

Many organizations when first dealing with HIPAA have exaggerated to be on the safe side and this has created hurdles for progress. When a new effort is made, it would be good to carefully consider a better balance to address the real requirements and not too much more.

Currently Shands and AHC-IT is managing EPIC and the IDR for storing PHI data. The IDR process allows for de-identification but its use is very limited. The most use currently from the IDR is the search for cohorts, which is a set of potential patients for a research study who have consented to be contacted for research.

In the age of big data, there is so much more that could be done if the data we have could be cross searched for trends. This is something of incredible potential to complex disciplines like the epidemiology department and the EPI. Other places like Massachusetts General and Harvard have been doing this for a while. UF needs to catch up.

Organization

The agenda for the summer 2013 can be found at http://www.it.ufl.edu/governance/advisorycommittees/researchcomputing.html.

Next meeting will be on July 1 at the usual location in NPB 2205 from 1:30 – 2: 30 pm.