September 25, 2018
Tuesday
2:15 – 3:30 pm
Speed Learning and Small Table Discussions

Next Generation Cyber-Systems and New Approaches for Data Intensive Sciences – Harvey Newman, CalTech/Tom Lehman, MAX/Edmund Yeh, Northeastern University (leads)

The following key issues and items were noted:
-R&E data is increasing at an exponential rate of 30-70% per year depending on which R&E network
-Network speed per unit cost is increasing at a much slower rate, closer to an average of 13% a year, based on technology upgrade cycles such as the 100G to 400G generation transition (of about 10 years)
-Inexpensive end systems can easily drive network traffic at rates equivalent to regional and wide-area network link speeds; these are getting less expensive year by year

The conclusions are that:
-There is an exponential mismatch between the data increase and the network speed increases.
-From a technology and cost perspective, it is easy to build a single end system which can overwhelm regional and wide area network paths
-The only reason that networks are ok today in terms of current utilization rates, is because i) site end systems networks are not well engineered and/or placed for performance, ii) it is still difficult to use and tune network and data movement systems.
-The fact that current R&E regional and wide area networks are not over utilized is really a negative reflection on current deployments and operations. Based on data volume increase and end system technologies/cost, most regional and wide area network links should be over utilized.
-If the current backbone links are not over utilized today that should considered a failure, which is due to the fact that networks/data movement systems are too difficult use and end sites/systems are not well engineered.

It is expected that these end system and network usage issues will get be resolved, and the result will be that network infrastructures will need more proactive management.

This has driven the development of intelligent software driven systems such as the one in the SENSE project (https://tinyurl.com/sense-project-overview) where networks and end-sites can be modeled, their topology can be understood and network resources can be allocated and managed, and optimized. These intelligent, stateful systems are able to react to multiple requests, negotiate and adapt.

The new data access paradigm NDN (Named Data Networking) (slides attached) provide a natural way to reference, locate and deliver data collections by name, and have new routing and caching algorithms that can accelerate data access and distribution. Generalizing this to edge computing - considering computing as well as data storage in caches, is an important trend, where relatively few algorithms and optimization approaches have yet been developed.

Bringing the SDN/SENSE and NDN threads of development could be a powerful new direction in the development of next generation systems.

There may be a need to balance between stateless and state-full network management to better utilize

network resources.

It was observed that a big problem with distributed management of cyberinfrastructure is the difficulty in obtaining state for distributed resources in a manner with end-to-end decisions can be made. New methods will be needed to allow resource owners to share just enough information to allow specific mutually beneficial outcomes to be achieved.

It was noted that now is the time to work on solutions based on constrained wide area network links, because once this happens it will happen rapidly.

Solutions based on edge computing/storage and hybrid cloud are promising technologies to provide better service to users, and better utilize cyberinfrastructure.

Named Data Networking (NDN) is another technology that should be evaluated in this context.

Automated network provisioning and management systems which allow for a range of usage modes is important. For some users continuing use of the network as a stateless entity will need to remain. For other users, a more interactive and deterministic network usage model may need to be developed. There are several projects such as the SENSE project which is working on this more interactive network model.

It was noted that network and cyberinfrastructure engineers with software development skills will be increasingly needed. Other skills that will be important are knowledge across a broad base of technology areas including networks, compute, storage. The human interaction component will also be critical to work with users to help them tailor their workflows to modern cyberinfrastructure.

The development cycle of distributed vs centralized cyberinfrastructure resources was noted. It seems like the current phase is revolving around a hybrid model where centralized cloud infrastructures and distributed on premise edge facilities will both be utilized in application specific ways.

It was noted that artificial intelligence, machine learning, and other intelligence focused technologies will be critical parts of future cyberinfrastructure.

It was noted that much of campus traffic is now Netflix.

A link to some additional slides with relevant information is provided here:
https://www.dropbox.com/s/vxtv8uxyjnaa4kl/NextGenCyberinfrastructureDIS_NSFCCStarMeet092518.pptx