2.13 IT utilities

2.13.1 Overview

Information technology is so fluid that we can only describe here what we would plan for IT utilities if we were building the ERL now. Extrapolation of storage capacity, CPU performance, and network speed can be taken into account, but accurate detailed plans are impossible at this point. What will actually be installed, as the ERL is constructed, will be based on the best technology available at the time and incorporate best practices gleaned from other accelerator facilities.

The CLASSE laboratory computing facility will provide the IT infrastructure for the ERL project. Although staffed and managed independently from the rest of the university, it exists within the context of the broader Cornell University IT structure. It uses Cornell's backbone network for connection to the Internet and university provided services, such as e-mail and scheduling. It is subject to university IT policies regarding network security, data stewardship, and other matters. The computing facility provides core services including those required for desktop support, network access, and data storage. The computing and network resources of the ERL Control System, described in §2.11, are an operationally autonomous part within the computing facility. Although there will be computing and data storage facilities for scientific analysis on site, the data processing and analysis of X-ray experiment data will be the responsibility of those doing an experiment and is expected to be addressed offsite. The Cornell Center for Advanced Computing (CAC) would be the natural facility for extensive experimental data analysis for scientists based at Cornell.

Physical infrastructure

The main computing facility consists of a server room in Wilson Laboratory. This is the network hub for distribution throughout the laboratory and connection to the campus network and the Internet. Adequate UPS power and $n + 1$ redundant cooling will be provided for the major server, storage, and network equipment of the laboratory and control system. As space is restricted at Wilson, equipment that does not need to be located directly on site, such as off-line computer clusters and backup storage, can be located across campus in the CLASSE server room in the Physical Sciences Building. The core superconducting RF (SRF) research and engineering facilities are also across campus in Newman Laboratory and will have computing equipment and a network connection via fiber-optic cable to Wilson Lab. Dedicated high-speed network connectivity can be provided between Wilson Lab. and the CAC, located approximately a kilometer away, as needed for experimental data analysis.

Basic services

The CLASSE laboratory computing facility will provide central support for all the basic IT services required for the laboratory. These include:

- Workstation hardware and software support
- System administration

- Centrally served file storage for individuals
- File system backup
- Software application licensing and support
- Videoconferencing support
- Collaboration tools
- Web services and support
- Service and issue tracking
- Business system support

The university's central IT organization, Cornell Information Technologies (CIT) provides basic services for CLASSE, and the ERL, including:

- E-mail
- Scheduling
- Software licensing for some applications
- Backbone Internet connectivity
- Network security services (e.g. intrusion monitoring)

Computer clusters

Although not intended for data analysis of ERL experiments, there will be one or more computer clusters to meet the computational needs of CLASSE and the ERL. For computation critical to ERL operations, including simulation and on-line modeling, a computer cluster will be located on site in Wilson Lab. For non-critical or more extensive computations, there will be a computer cluster located in the Physical Sciences Building with a batch queuing system for efficient shared use.

University research computing facility

For computations beyond the capacity of the laboratory facility, CAC provides high-performance computing systems, data storage, and applications to researchers at Cornell and their collaborators. The CAC serves faculty researchers from dozens of disciplines and is a natural choice for Cornell-affiliated ERL users with significant computing requirements. CAC staff has extensive expertise in HPC systems, storage, database systems, data analysis, and visualization. CAC network connectivity includes TeraGrid, National Lambda Rail, Internet2, and New York State Grid. As a core facility of Cornell University, CAC receives funding from Cornell and its supporters, including the National Science Foundation. CAC's financial model is designed to recover the cost of operation, upgrade, and replacement of the equipment used.

Storage

Extensive storage capability is available to meet ERL requirements. RAID-based disk storage on a storage area network can be extended to the Petabyte scale as needed for simulations, machine physics data, ERL control system archives, and experimental data if required. Long-term storage on tape is also available.

Network

The ERL control system necessitates heavy use of network segmentation and throughput. Centrally located core switches with level-three routing capability provide the basic partitioning and connectivity of the backbone network. These isolate major traffic flows and separate the control system network from general purpose laboratory network functions for uninterrupted control operation.

The backbone network is based on 10 Gbit Ethernet technology, with 10 Gbit trunk fibers radiating out from central switches to local front-end switches connecting 1 Gbit Ethernet to equipment shelves and isolated pieces of equipment. Multiple trunks are used where needed to provide adequate throughput/latency, separating types of traffic where appropriate. The Ethernet network is modular enough to permit easily implemented adaptation over time to traffic patterns and potential bottlenecks. The ERL will have wired and wireless network access deployed throughout the entire facility, for general purpose and maintenance purposes. Segmentation into appropriate subnets will provide protection for the control system as well as remote management of control system and ERL components. In experimental areas, separate wired and wireless network access will be available for the needs of outside users while maintaining suitable isolation from other laboratory networks.

Access and security

The overall plan for access and security is to provide robust authentication for everyone within the facility and to establish appropriate authorization for access based on users' roles and needs. Routers and a central firewall will isolate and protect the ERL controls network from the general laboratory network and from the Internet. At the same time, they will selectively allow expert entry from the outside and access to remote resources across the boundary, as needed, in a carefully restricted and monitored fashion. Logging, intrusion detection and careful monitoring will be deployed to maintain security of all laboratory subnets, with special attention to the controls network. Special provisions on a separate network will be made for outside experiment staff and visitors to have access appropriate to their needs without undue restrictions.

2.13.2 Electronic document management system

CLASSE currently uses Invenio, from CERN, running on a local server, against a local backend MySQL database, for its general document EDMS needs. We plan to expand and maintain this to meet the document needs of ERL design, construction, and operations. CLASSE uses Autodesk Inventor for mechanical drawings with Vault as a repository.

Business systems IT facilities

Since CLASSE is a unit of Cornell University, the CLASSE Business Center, responsible for ERL business systems, operates within the broader context of Cornell's central IT systems. Besides local workstations, all business IT facilities are supplied and supported by Cornell University.

Relational data base

A relational database facility (currently MySQL) will be provided by the laboratory to meet the database needs of the control system, business center, as well as the miscellaneous requirements of running a scientific laboratory, such as documentation, record keeping, parts inventories, and personnel information. Appropriate provisions will be made for replication, backup, segmentation, performance, availability, and isolation, as needed.

2.13.3 ERL control system needs

The computing and network needs of the ERL control system are described in §2.11. Although logically within the context of the CLASSE laboratory computing facility, the ERL-controls-IT infrastructure will be largely autonomous, such that controls-T availability will not be affected by interruptions of service, planned or unplanned, of the laboratory facility or the university. Because they are so critical to the ERL, laboratory utilities such as power and cooling will be monitored and, where appropriate, controlled, as part of the ERL control system.

2.13.4 Beamline experiment IT needs

In the new era of pixel detectors, the potential for data rates from X-ray experiments is enormous. At full rate, not only would network and storage capacities be exceeded, but processing all the data off-line or after the experiment finished would be impractical. Clearly, methods of drastically reducing the data from these detectors must be developed in the intervening years before the ERL is operational. It is now too early in this process to make realistic plans for needed IT resources. With this assumption, the experimental collaborations must have primary responsibility for data acquisition, on-line data reduction, storage, processing, and analysis of X-ray experiment data. At the same time, there will be on-site capacity for experiments with low to modest data rates and for experiments that choose to move reasonable amounts of data (e.g. Gbytes to a few Tbytes) in real time to off-site locations. Cornell-affiliated experiment collaborations may choose to process data at the CAC, and for this, we will have 10 Gbit network connectivity to address their needs. Most other experiments are expected to provide their own local data acquisition, processing, and storage capabilities, transporting their resulting data for further analysis at home institutions via portable storage devices.