Distributed Computing Environment in the ESnet Community

R. Roy Whitney

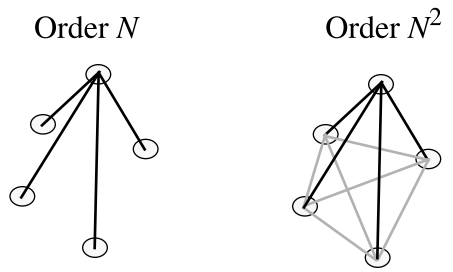
Continuous Electron Beam Accelerator Facility

12000 Jefferson Avenue

Newport News, VA 23601

whitney@cebaf.gov, https://www.cebaf.gov

The Energy Sciences Network (ESnet) Community is undertaking an ambitious project to coordinate its distributed computing. A reasonable question is: *"Why Coordinate Distributed Computing?"* The answer can be seen in the figure below:



The issue is complexity.

Without a coordinating element, the number of links for communicating between a number of nodes would grow with order N^2. Inserting an effective coordinating principle can reduce the number of independently managed links to order N while still leaving all of the links available for actual communications. A good example from telecommunications is the public phone system where we rely on central coordination of the phone numbers rather than having to personally negotiate the switching information with everyone with whom we wish to communicate. The Internet's distributed system of name servers is another good example.

The meta-question to the first question is: *"Why use distributed computing?"* The answer can be put in three parts:

- Individual end users require access to far more information and computational resources than they can acquire locally.
- For some applications, distributed computers are nearly as efficient as centralized computers and the total CPU resources available may be significant.
- Many resources are unique. They cannot easily be replicated. Some examples in the DOE community include research accelerators, fusion facilities, systems of virtual caves, and environmental laboratories.

The overall response to the questions can be summarized in the concept of virtual laboratories. Humans want their resources responding within a second and at arm's length or closer. Using coordinated distributed computing, these requirements can be met by placing facilities on-line via Internet style technologies. Further, human communication is 10% verbal, 40% audio and 50% physiological. Using combinations of words, sounds and pictures including animation, virtual laboratory technologies can exploit all of these aspects of human communication. The human end user receives the benefit of virtual presence at on-line facilities and makes use of a rich collection of resources.

The ESnet Community has three main sources for its distributed computing vision, requirements and technological capabilities:

- End users
- U.S. Department of Energy (DOE) /Energy Research (ER) /Office of Computing and Technology Resources (OCTR) /Mathematics, Information and Computing Sciences Division (MICS)
- ESnet Steering Committee (ESSC)
  - ESnet Site Coordinating Committee (ESCC)
  - Distributed Computing Coordination Committee (DCCC)

The end users, DOE/OCTR/MICS and ESSC focus on the vision and requirements. The ESCC and DCCC focus on the technological capabilities and implementation.

The current ESnet distributed computing project is Distributed Informatics, Computing & Collaborative Environment (DICCE). The goals for this project are to:

- Set up DICCE for the ESnet Community
- Create the basis for:
  - Virtual Laboratories
  - Networked Collaborative Environments
  - Facilities On-line
  - Distributed Computing Environments

The minimum DICCE is:

- High quality Wide Area Network (WAN)
- Key distribution system
  - Kerberos and public
- Authentication and authorization services
- Open Software Foundation's Distributed Computing Environment and Distributed File System (OSF's DCE & DFS)
- Secure MIME compliant E-mail
- Secure WWW technologies

While most of the funding for DICCE comes directly from the participating facilities, in 1995 DOE/OCTR/MICS provided some funding for a DCCC coordinated set of proposals for DICCE to be used for Research and Development (R&D) activities. More information can be found on these proposals under https://www.es.net. DICCE will be a multi-year project as it builds infrastructure essential for the virtual laboratory complex.

The first requirement for meeting the minimum DICCE is to have a high quality network. ESnet provides T3 (45 megabits/sec) Internet-style connectivity to all of the major DOE/ER laboratories, a variety of other DOE sites, and several DOE centers at universities. Additionally, ESnet provides these facilities with connectivity to the Global Internet via connections to the major commercial Network Service Providers and many regional networks. ESnet was the first major backbone piece of the Global Internet to provide these services using Asychronous Transfer Mode (ATM) technologies. Some portions of ESnet will be upgraded to OC3 (155 megabits/sec) in the near future. ESnet is managed by Jim Leighton at the National Energy Research Supercomputer Center (NERSC). Currently ESnet is moving with NERSC from Lawrence Livermore National Laboratory to Lawrence Berkeley National Laboratory.

The overall DICCE project is managed and coordinated through the ESCC and DCCC via task forces and working groups. The following is a listing of the DICCE-related ESCC and DCCC groups along with information on the ESSC:

- ESnet Steering Committee (ESSC) — Sandy Merola (LBNL) Chair
  Provide programmatic goals, vision and user input to ESnet management.
- ESnet Site Coordinating Committee (ESCC) — Roy Whitney (CEBAF) Chair
  Coordinate site issues for the effective implementation of ESnet at the sites.
- Distributed Computing Coordinating Committee (ESCC) — Roy Whitney (CEBAF) Chair
  Coordinate the implementation of ESnet Community distributed computing. Provide project management for DICCE.
- Key Distribution TF — Bill Johnston (LBNL) Chair
  Set up key distribution for Kerberos V5, OSF/DCE and Public Keys.
- Authentication TF — Doug Engert (ANL) Chair
  Set up cross realm Kerberos V5 authentication ESnet Community wide.
- IPng WG — Bob Fink (LBNL) Chair

Set up IPng open test bed. IPng is critical to long-term DICCE goals when extended to National and Global Information Infrastructure activities.

- Network Monitoring TF — Les Cottrell (SLAC) Chair
  Provide tools for LAN and WAN monitoring.
- E-mail TF — Mark Rosenberg (LBNL) Chair
  Institute MIME, PEM and PGP compliant E-mail ESnet wide. This technology will also serve as the basis for digital signatures.
- Remote Conferencing WG — Kipp Kippenhan (FNAL) Chair
  Advance collaborative video conferencing.
- Distributed Computing Environment WG — Barry Howard (NERSC) Chair
  Coordinate implementation of OSF's DCE, CORBA, etc.
- Andrew/Distributed File System TF — Troy Thompson (PNNL) Chair
  Set up an ESnet Community-wide file system.
- Distributed Systems Management WG — John Volmer (ANL) Chair
  Provide structure for implementing systems management in the ESnet DICCE.
- Applications WG — Dick Kouzes (WVU) Chair
  Coordinate DICCE applications development including underlying tools.
- Group Communications WG — Allen Sturtevant (NERSC) Chair
  Coordinate file types supported in the ESnet Community for ftp servers, WWW, MIME, etc.
- Architecture TF — Arthurine Breckenridge (SNL) Chair
  Consider high-level issues for setting up DICCE throughout the ESnet Community.

Clearly a significant level of resources isbeing focused on the distributed computing infrastructure at the DOE facilities. It is appropriate to comment on one project of particular interest to the ICALEPCS community. The DCCC DICCE proposal was actually made up of 23 separate proposals. DOE/OCTR/MICS approved nine of these proposals and funded seven. The 23 proposals were self-rated by the DCCC proponents. The number two rated proposal was that from the Experimental Physics and Industrial Controls System (EPICS) collaboration to put a DICCE layer around EPICS so that facilities running EPICS could be brought on-line via the Internet in a secure fashion.

As previously noted, one of the DICCE goals is to provide a global file system available to the entire ESnet Community. Presently, Argonne National Laboratory (ANL), NERSC/LLNL, Pacific Northwest National Laboratory (PNNL), and Sandia National Laboratory (SNL) have OSF/DCE including DFS on-line in a cross-realm configuration, i.e. a simple change directory command moves the user from the file system at one facility to another. Ames Laboratory, CEBAF, LBNL and several universities will also soon join the file system. The goal is to move this DICCE pilot to a production service for the ESnet Community distributed collaborations by Summer 1996. Note that meeting this goal implies that key distribution systems have been made operational, a significant challenge.

Higher level services are anticipated by Fall 1996. Long-term DICCE milestones include providing a "Single Environment" for ESnet Community collaboratory end users, a diverse set of virtual laboratory tools and applications, and all ESnet Community facilities on-line where applicable.

Additionally, since the start of the DCCC DICCE project the World Wide Web has switched from being driven by R&D interests to being driven by commercial interests. For example, the technologies for authentication and authorization on the Web are being determined by banks and credit card companies. As the Web is proving a versatile virtual laboratory tool, the DICCE project is busily incorporating these Web offerings. The end user will require only one Web browser to securely interface with and control on-line facilities, data, and other budget, personnel and management information.

Consider this prediction: Your successful R&D laboratory will aggressively deploy mission-oriented virtual laboratory/DICCE technologies; complexity will be coordinated and your overall organization will have a higher productivity. As you visit different facilities or communicate with your colleagues, carefully check what is happening with network access to the facilities, scientific and engineering data and information, and management information system resources. Check to see and hear how the more successful projects are achieving their goals.

In summary, because ESnet is a mission-oriented network, its community can directly attack the R&D challenge of setting up the distributed computing infrastructure for virtual laboratories in a coordinated manner. Its scientists and engineers work together to meet their programmatic requirements. Many university researchers are also involved. It is anticipated that many of these ESnet distributed computing projects will have significant impact far beyond just the DOE environment — reminiscent of how the European Center for Nuclear Research's (CERN's) World Wide Web has impacted society. Distributed computing and virtual laboratories will be essential elements of the National and Global Information Infrastructure.