HEPiX Minutes
March 24–26, 1993
CEBAF

The third meeting of HEPiX–US was held March 24–26, 1993 at CEBAF in Newport News, VA. Twenty–one members from nine different sites were in attendance. The following is an attendee list:

Name	Affiliation	E–Mail Address
Bisson, Ernie	MIT	bisson@sue.mit.edu
Burton, Jackie	SSCL	burton@ssc.gov
Chambers, Rita	CEBAF	chambers@cebaf.gov
Cormell, Larry	SSCL	cormell@ssc.gov
Cottrell, Les	SLAC	cottrell@slacvm.slac.stanford.edu
Dassonville, John	SSCL	dassonville@ssc.gov
Giacchetti, Lisa	Fermi	lisa@dcdlaa.fnal.gov
Koenen, Frank	Fermi	koenen@cdsun2.fnal.gov
Lauer, Rochelle	Yale	lauer%yalph2.bitnet@yalevm.ycc.yale.edu
Liang, Gebi	CEBAF	gebi@cebaf.gov
Love, William	BNL	love@bnl.gov
Mastroserio, Paolo	INFN	mastroserio@vaxna1.na.infn.it
Mc Fadden, Ed	BNL	emc@ax61.bnl.gov
Mylonas, Rudolf	Paul Scherrer Institut	mylonas@csun.psi.ch
Nicholls, Judy	Fermi	nicholls@fnal.fnal.gov
Philpott, Sandy	CEBAF	philpott@cebaf.gov
Scipioni, Brian	SSCL	brian@ssc.gov
Selover, Mark	SSCL	selover@ssc.gov
Sullivan, Kevin	Fermi	kevins@baja.fnal.gov
Watson, Chip	CEBAF	watson@cebaf.gov
Whitney, Roy	CEBAF	whitney@cebaf.gov

Wednesday, 3/24

Sandy Philpott (CEBAF) opened the meeting Wednesday morning, followed by Roy Whitney's (CEBAF) overview of the CEBAF facility. Judy Nicholls (Fermilab) spoke to the group about HEPiX topics, including where the idea of HEPiX came from, the 2nd HEPiX meeting at CERN, why we gathered (to discuss ideas and exchange our knowledge on UNIX issues: different flavors of UNIX; open systems; distributed computing..), and what we hoped to accomplish at the meeting.

Site reports followed, with each site giving an overview of systems and projects of interest to the HEPiX community at their home institution.

BNL (Computing) – Ed McFadden
Hardware Support:
Install workstations, PCs, Macs, and communication facilities like FDDI, Ethernet, etc. Most of the workstations are Suns; SGIs are second; then RS6000 followed by DECStations. HPs are on site but not supported. For hardware maintenance service, labor is prepaid for problem diagnosis and parts replacement is paid by user. There are also workstations maintained by vendors.
Platform Support:
UNIX (SUN, IBM, SGI). Total staff 8–9. Install the operating systems for new machines. Make the machine ready to use and maintain the systems.
Office Technology Center (Office Automation):
Support PCs and MACs. Provide the configuration support and applications like E–Mail, CAD/CAM.
System Support Services Section:
The System Support Service Section has downsized from 2 staffs to 2 operators. The computer farm (DNQS) includes AIX, Alpha, SUN, SGI. Interactive UNIX systems include ULTRIX, AIX, ALPHA OSF. Fileserver is a SUN SPARC. ABARS provides Backup, Restore and Archive functions. VMS cluster provides interactive logins and batch queues. UNIX and VMS software is supported including 3rd party and public domain software.

BNL (Physics and RHIC) − Bill Love
Project STAR's computing choices: abandon VMS, include HP−UX, IRIX, AIX and SUNOS followed by ALPHA OSF, Fortran90 and C, CVS−RCVS code management. Data structures kept in commercial database.

Project STAR's computing problems: RCVS is new; Event Simulation files are large(1. Scratch Disk Management strained. 2. Need good tape management); Collaboration is scattered.

CEBAF − Sandy Philpott
Computer Center is VMS and UNIX mixed; support two platforms: DEC and HP. The central VMS systems are three Vaxstation 4000/90s and an 8700. The main UNIX computers are a DEC5000/200 running Ultrix (plus a 5000/240 for Ingres), and an HP9000/730 (soon to be 735). Two staff members manage the centrally managed UNIX computers: approx. 30 DECstations, 40 HPs, and 20 Tektronix/NCD X terminals. Questions included whether X on PCs and MACs are supported−−at this time CEBAF supports X on MACs but not on PCs as it's still quite complicated.

Fermilab − Judy Nicholls
No direct support for PCs and MACs. A main consulting number is provided. Network growth has been tremendous. Currently there are 4000 nodes on the network. CPU growth is large too. Most users use VMS system and UNIX systems. In 1993, UNIX systems supported are 600. 8 personnel are assigned. The platforms supported are IBM, SUN, DEC and SGI. WWW is implemented to pull both VMS and UNIX information. Judy's report began the discussion of WAIS versus WWW.
INFN − Paolo Mastroserio
Computers include ALPHA, VAX, DECStations, HP, SUN, X terminals. A system package called DECnsr (DEC network save and restore) is developed to ease the software task in a multi−vendor environment. Backup package can write and read label to/from tape.

Paul Scherrer Institut − Rudolf Mylonas
PSI Computing Division has 32 members with 3 subdivisions: System Management and Operation(12), Communication (7) and Project Services(10). Support is limited to 2 platforms: VMS and UNIX. The main computing facilities are VAX cluster and CONVEX. They are used about 90% in batch mode. Besides,there are 67 Sun's, 10 Dec stations and 25 IBM RISC stations. Software support includes DecWrite, TEX/LATEX, VNI, AVS, IMSL, CERN libraries, Maple, Matlab, Archie, etc. Information services include VAX notes, Electronice 'telephone book', DEC BookReader, Sun AnswerBook, etc. Communication services include X windowing,Mail, Ethernet and FDDI, 9 track, 1/4''cartridge tapes and 8 mm Exabytes and distributed printing. Future plans include seting up farms that satisfy CPU needs and Unique mail system, etc.

SLAC − Les Cottrell
Environment today: Workstation: PCs(500), MAC(400), VMS(140), UNIX(200), X terminals(120). NFS running on UNIX, VM, VMS and PC. Single YP domain, account aligned across centrally managed machines. Applications include Rexx, Perl, Xapps(Frame, etc), WWW, E−mail(X.400), Netnews, whois, finger, Oracle, WDSF−>DF/DMSfor archive/backup. There is a UNIX Coordinating Committee, UNIX Journal Club. Staffing for UNIX is going up while VMS is going down.

Model for Future Computing at SLAC: Major services will move to UNIX (compute and data). Majority of users are on PCs & MACs and are expected to stay as better productivity and lower cost can be achieved. Specialized users use X terminals with a dedicated support host. UNIX workstations for those with "power" needs (Sun, IBM RS/6000).

Major Focus Areas: Improve management of distributed environment: Reduce administrative costs; provide acounting and resource management; improve availability. Provide & manage mass storage. Real time UNIX with AIX. Improve connectivity to home. Improve network management.

Following the site reports, sheduled talks were presented to the group.

'ESnet Site Coordinating Committee Activities', and

‘Energy Research Distributed Computing Environment Activities’, Roy Whitney, CEBAF

Roy talked about what is going on at Enery Science Site Coordinating Committee (ESCC) and Enery Research Distributed Computing Working Group. ESnet DECnet Working Group (EDWG) has worked out some DECnet Phase V transition documents, which are available through anonymous ftp at ESnet NIC. ESCC’s Remote Conerencing Working Group will have Video Control Center soon to be at NERSC for the Engy Research Video Network. The transition plan will be in place by summer. ESCC’s Desktop Video Task Force facilitates a pilot project in the ESnet environment using the desktop video technology being developed at LBL. This group evaluates disktop video options for implementation by Enery Research groups on the ESnet and the associated Internet. The Foreign Connectiviy Issues discussed at ESCC include: EBONE, EMPB, Connectivity to Pacific Rim and Other Areas Growing, CIS/FSU/COCOMP Connectivity document, and ESCC Russian Federation Task Force. ESCC’s AFS task force has the final report available through anonymous ftp. The Security Working Group & Authentication Task Force has several pilot projects but none of the authentication systems are ready for general WAN usage. The X.400/X.500 Task Force’s goal for 1993 is to have all ER sites on−line with wp.site.edu/gov format and to expand the number of sites and integrate with advanced smtp technologies and incorporate MIMI.

The High Speed LAN/WAN Workshop held at Berkeley last November was reported. The goal of the workshop is to develop stategies for solving LAN issues so the ESnet sites will be in positions to make effective use fo the ESnet WAN capabilites as they evolve. Topics include: ESnet ATM WAN Project, ATM LAN project, HPPI, Fiber Channel, FDDI, SCI. Distributed Computing Working Groups Ideas are introduced. They include: Distributed Computing (DC), Distributed File Systems (DFS), Distributed Computing Environment & Management (DCE&M), Distributed Database Systems (DDbS), and Distributed Mass Storage Systems (DMSS).

‘Fermi’s Experiences and Plans for AFS’, Lisa Giacchetti, Fermilab

A committee was established more than 1 year ago to investigate AFS(Andrew File System) and determine if Fermi should implement AFS into any of its computing projects. The committe found both advantages and disadvantages and decieded to: (a) Implement AFS on FNALU, a new project that whould provide UNIX batch and interactive computing (b) Use Transarc Corporation’s version of AFS since they had a version of AFS running under AIX and were developing a port for IRIX.

There are certain constraints influenced the organization of products on FNALU. For example, products needed to be available from within AFS. So they decide the product area would be setup in a special way to meet the constrains. One example of the rules is that the file tree would look like other multi−flavor products areas that had been setup on site with one exception − it would start at /afs/fnal/products not /usr/products.

FANLU and FERMI’s AFS cell will be completely FUE’ized but the transmission will be slow. The parts of FUE that are known to work include: the UPS and UPD products support facilities; standard FERMI login shell scripts, etc.

FNALU is still in its infancy. It currently has two IBM 350’s dedicated to file serving, one IBM 560 dedicated to interactive user and one IBM 560 dedicated to batch use. The Computing division are still working on establishing a products base and porting FUE to AFS.

Issues need to be resolved: (a) user education (b) training for system managers (c) rewrite/develop system management utilities (d) rsh/rlogin /cp triad of terror (e) batch system integration

Future plans for AFS: (a) data file serving (b) serving of software products and executables (c) distribution of OS software.

'OSF on the ALPHA−AXP', Rochelle Lauer, Yale High Energy Physics
ALPHA−AXP's performance is quite satisfying. Yale has done some testing on ALPHA (runing GENANT). Some benchmarks are reported.

From their experience, it's hard to tell what OSF is. There are no documentation yet. It's typical UNIX. Vendor additions merge with OSF functions. It's no surprise that it's a combination of BSD and AT&T. New kernel is clean and neat. One can search out vendor specifics by looking at the headings in man pages and scripts. For system management, many scripts control startup and shutdown. Scripts are controlled by environment variables. Vendor supplies setup scritps like netsetup. There is no DME. File system includes UFS and NFS. For migration, ALPHA OSF is not binary compatible with other UNIX. ALPHA VMS is binary compatible. Some (un)usual surprises include: #!/sbin/csh executed from other shells; YP has problems, etc.

Overall, OSF is a better UNIX with pluses like common kernel, common management methodology (e.g. startup, crontab etc) and after all it's all purpose UNIX user environment. The minuses about it are: diverse vendor supplied tools; no DME; no DCE; limited migration path, etc.

In conlusion, OSF is just another UNIX. End−User incompatible problem is mostly due to hardware architecture. There are still management issues in multi−vendor distributed environment.

'PDSF (Physics Detector Simulation Facility) Update and Plans', Brian Scipioni, SSC
PDSF(Physics Detector Simulation Facility) by Phase I (3/91) had CPU Power(MIPS) 2000, On−line Storage(GB) 50 and Tertiary Storage (TB) 0.25. By Phyase II (4/92), it has CPU Powers(MIPS) 2000, On−line Storage(GB) 150 and Tertiary Storage(TB) 0.25. CAWG I (Computer Acquisition Working Group) was formed in early '90 to plan, design and procure Phase I of the PDSF. CAWG II was formed in summer '91 to plan, design, and procure Phase II of the PDSF. Network layouts for both Phases are shown. PDSF utilization history report is given between 5/14/92 to 10/08/92. Daily average CPU utilization for different systems ranges from 30% to almost 100%. PDSF has a total user no of 406.

Some PDSF phase III considerations include: (a) 4000 SSCUPS @ 200 GBs (b) Separate Compute service and fileservice (c) Batch resource is main focus (d) Architecture supports workgroup concept (e) Current architecture scales.

A proposed PDSF III hardware configuration is shown.

'Farm activity at Brookhaven using DNQS', Ed McFadden, BNL
Some acutal batch job running output was presented and discussed.

'Experience with UNITREE', Rudolf Mylonas, Paul Scherrer Institut
A survey in 1991 was done to find out what users need. The result indicated an approximately 350 GB system needed to be accessible from both VMS & UNIX worlds by about equal amounts. UniTree was selected for the following reasons: (a) a virtual disk system (b) hardware independent (c) variety of media are supported. The installation on the convex took place in December 1992. Beforehand, convex's memory was upgraded to 256MB and the FDDI connection was made between VAXcluster and the convex. The system is in a test phase until the end of March, 1993. During the test period, it's open to users on a no guarantee basis.

Access to UniTree is currently only using FTP since convex NFS & UniTree NFS are not compatible now. As far as performance is concerned, file transfer rates is 1.2 Mbytes/sec from convex. Tape writes 9 min/gb.

Tape load is fast and tape mount takes 45 sec. Performance is bad for consecutive small files due to File Maker problem. Automatic file migration is done based on high & low water marking policy. No significant impact on CPU and memory.

From the user's point of view, UniTree works better with large files than many small ones. File is clearly marked if "archived". User is kept up−to−date every 15 sec during retrieval. From the system manager's point of view, time and effort need to be put to know the system limits & behaviour and additional account management is required as each user needs not only a convex account, but also UniTree access.

Overall, test users are positive about it. The access is 4 times faster than to a WORM disk. There is no noticeable degradation in the performance, operation or stability of the convex. FDDI response is not optimal at the moment.

After the meeting closed on Wednesday, the HEPiX dinner was held at Bon Apetit, a local French/Vietnamese restaurant.

Thursday, 3/25

Scheduled talks continued:

'Plans for Distributed Applications File Service and Common User Environment at SSC', Mark Selover, SSC

SSC needs new type of File Service for common software applications for the following reasons: (a) SSC has a large UNIX environment but no uniformity outside PDSF (b) Non−detector applications need to be distributed outside off PDSF (c) Importand to share the work load and to make efficient use of the software experts in SDC and GEM, etc. The plans are (a) build SSC distributed file service for common software (b) common software file service must be installed and tested before developing common user environment (c) develop basic SSC common user environment (d) evolve SSC environment with HEPix standards when appropriate. What is happening now are: (a) building and testing common file service for HP−UX, IBM AIX, SUN−OS, DEC Ultrix systems (b) SDC and GEM both agreed to install software distributions, directory trees in sscfs (c) basic framework for ssc login scripts has been agreed to by SDC, GEM, IS. They are building now (a) distributed file service is a NFS/AFS hybrid system (b) AFS provides transport between servers distributed over wide area network. (c) sscfs exported as read only service by translators (d) software developers and maintainers work in AFS (e) the common file path used on all client systems is /usr/ssc, a link on NFS clients to automount file system /sscfs/machine−type and a link on AFS clients to /afs/ssc.gov (f) All software is installed with the /usr/ssc as the root path (g) primary servers for sscfs will be the translator machines. (h) the current test system consists of 4 AFS volume/file servers with 30 Gbytes disk total and 4 NFS/AFS translators with 100 MB disk cache each. Users see same file structure for sscfs everywhere. Software maintenance is on 1 common file base. Login scripts based on software in sscfs can be standardized and put in common area.

'CEBAF's UNIX Systems Environment', Sandy Philpott, CEBAF

Since CEBAF is a relatively young lab, the computing environment has had controlled growth from the early days of VMS into DEC Ultrix, and is now moving toward HP−UX on the HP 9000/700 platform. Site UNIX management strategies include clustering, use of local disks, standard site directories (/usr/siten, /usr/local, /usr/usern, /<node>/usr/local,/<node>/usr/users...), setup script, and standardized system files.

Ongoing issues include NIS,YP between DEC and HP systems, standardizing shells, implementation of Automounter, easy software

distribution and update, and help and information retrieval. (CEBAF has just recently begun using WWW.)

Future directions to be decided are OSF, the Alpha architecture, and PC/MAC integration.

Major note was that VMS in NOT going away! Current VMS purchases are VAXstation 4000/90s that can easily be upgraded to Alphas running OpenVMS if desired.

'Experiences at Fermilab with Heterogeneous Cluster Computing on UNIX Workstations', Kevin Sullivan, Fermilab

Computational challenges include large (month−long) tasks and I/O, dynamic production environment and hardware configs, and round−the−clock support. CPS, POSIX−compliant production software, is very portable and, combined with clustered computing and standard LAN configuration, addresses these issues. I/O involves typically terabytes of data; 20,000 tapes; 200 Exabyte tapes per day. Fermi uses xoper tape mount facility; standard interfaces to uses and tapes. Dynamic production environment requires constantly adding and removing users, tape drives, disks, and cpu allocation, all of which are done without disturbing running jobs.

'SLAC Batch Computing Plans', Les Cottrell, SLAC

SLAC is looking to replace traditional mainframes with RISC clusters, and is evaluating BQS, DQS,and Load Leveller for batch software. Their experience is mostly with Load Leveller, and have found that it works better than the mainframes, and that multistreaming mechanism can be adapted to other programs. SLAC is experimenting with dedicated clusters on FDDI, using existing STK Silos, and is looking at IBM's SP1 for scalable parallel processing.

'Tools Database', Ernie Bisson, MIT Lab for Nuclear Science

Ernie asked for input on a tools database to provide an easy method for HEP sites to obtain commonly used Physics tools, to be made available via WWW and FreeHEP. For HEPiX entries, Ernie could be the FreeHEP editor, and offered MIT as a software storage server.

'UNIX Issues in the CEBAF Data Acquisition System', Chip Watson, CEBAF

Chip Watson, head of the CEBAF Data Aquisition group, discussed CEBAF's upcoming computing needs. 19mm looks to be a promising recording technology for data storage, with 60TB in a 2−axis robot, because it meets the 8MB/sec transfer rate. Other technologies discussed included 2480, 8mm, VLDS, and OPT. Networking discussions included FDDI, Fibre Channel, and ATM, as well as the need for standard lightweight protocols (processing now takes 6−8 MIPS for 1MB/sec). UNIX issues from a data acquisition developer included the suggestion to list all known sockets, nodes, ports, and program numbers to remove conflicts−−a location broker?; security issues−−programmable Ethernet addresses, outside access needs for experiment control vs. no outside access to accelerator control system, etc. and network impact of the data acquisition system −− using 100% bandwidth, broadcasting, and heavy X−window usage.

'LUE: A simple tool for distributed management and resource tracking', Frank Koenen, Fermilab

Fermi's Local User Environment (LUE) is a simple mechanism using electronic mail to collect information about computer systems: administrative, hardware config, OS, network ID, etc. Frank presented the security issues and resolutions encountered during the LUE implementation, as well as sample usage.

'Enterprise−wide Drawing Management Issues and Solutions', Frank Koenen, Fermilab

DCS, the Document Control System developed and currently in use at Fermi, manages CAD data to provide working, released, and archived

drawing tracking and control file revision. It's relatively inexpensive custom development meets all of Fermi's needs, and interfaces to other applications, including Oracle and Motif.

'Software Support in a Distributed Computing Environment: An Update', Judith Nicholls, Fermilab

Fermi's ups (UNIX Product Support) and upd (UNIX Product Distribute) are still the basic methods of software support and distribution; new products have been added for the UNIX and VMS platforms. upp (UNIX Product Pull) automates the process of obtaining software on UNIX (pulling when desired instead of having software pushed when it isn't wanted or there isn't enough disk space); vpp performs the same function for VMS. It's not clear at this point what part AFS will play in distributing software.

'Site Administration',John Dassonville, SCCL

Computer Operations are subdivided into the following groups: System Operators, System Support(VMS), Systems Support(UNIX), Systems Support (Micros), Info Center, Hardware Support. System Operations has 5 staffs. It is responsible for daily operations of central resources (backup, account management, trouble shooting log, etc.). VMS System Support has 2 staffs. It is responsible for VAX/ VMS system management and system programming of the information management and technical VAX clusters. UNIX System Support is responsible for system management and programming of both the centralized and distributed UNIX computing resouces. Microsystem Support has 4 staffs. It is responsible for system management and programming for microsystems including NOVELL, MAC OS and DOS. Information Center has 6 staffs. It is the focal point for user support. Hardware Support provides/coordinates hardware repair of desktop systems.

'CLUBS', Judy Nicholls, Fermilab

Judy presented the Clustered Large UNIX Batch Systems environment at Fermi. They need 1000 VUPs of high performance batch with high I/O capability. Initial config is Amdahl data server and RISC compute servers (RS6000); future enhancements include SCI support and RISC data servers, Unitree, CLUBS AFS environment, STK Silos or new robotics.

Friday, 3/26

The meeting closed Friday with discussions between the ten remaining attendees, plus an additional talk.

The group provides the following in reply to Roy Whitney's request for a HEPiX statement to the ER DCWG:

.

"We, the HEPiX−US group, support the formation of an Energy Research (ER) Distributed Computing Working Group (DCWG). We recognize the issues of distributed computing are very important to the ER program and in particular to High Energy & Nuclear Physics.

We believe that HEPiX−US, due to our extensive experience and in some cases leadership in the distributed computing arena, can provide important guidance and contributions to the ER DCWG activities. We therefore encourage the close interfacing between HEPiX−US and the evolving ER DCWG efforts."

'C/Fortran InterfaceProblems', Les Cottrell, SLAC

Les presented C<−−>Fortran interface problems to the group: 1. different calling sequences by different vendors that make it impossible to use the same C routine with Fortran on varied platforms, and 2. conflicting name spaces in libraries.

Program number registration

Discussion continued on the issues brought up by Chip Watson on program and version number registration. Is it a DCWG problem? Could

HEP get an assigned number group from SUN? What else needs registering?

Next Meeting:

The next meeting is tentatively to be hosted by Ed McFadden at Brookhaven National Laboratory (BNL) in the August−September time frame (he will verify when he returns to BNL). The backup site is SLAC, with host Les Cottrell.

The group encourages people to respond to the batch questionnaire and to post information in the newsgroup about their activities.

The group chose three issues for the focus of the next meeting:

Mass Storage
Batch
Security

Suggestions for the future included:

trying to involve more end−users of the HEPiX computer systems; including future directions of their sites in site reports; posting site activities to the news group or mailing list at the half−way point between meetings.

The meeting officially adjourned at lunchtime. Four members remained for the afternoon tour of the CEBAF Accelerator Site provided by Mike Syptak of CEBAF’s Physics division.