HEPiX Nikhef Meeting Minutes

Alan Silverman

July 2, 1993

These are the minutes of the HEPiX Europe meeting held on April 19th and 20th in NIKHEF, Amsterdam. Overheads of the speakers are available for most sessions, either from the author of the session; or via WWW via the route HEP - NIKHEF - MEETINGS - HEPMIX talks or the route HEP - HEPIX - NIKHEF talks; or by using anonymous to ftp.nikhef.nl and looking in the directory /pub/hepmix.

The meeting was held jointly with HEPVMx and attracted 66 attendees from 25 European sites plus the Coordinator of the North American chapter of HEPiX. The audience was predominately system support staff as opposed to users. Thanks for hosting the meeting and for the smooth organisation are due to the NIKHEF authorities, especially K.Gaemers, and to those individuals who devoted their time and efforts, particularly Willem van Leeuwen and Eric Wassenaar for the meeting arrangements, Rob Blokzijl who organised the worldwide audio conferencing of the sessions and the ladies in the secretariat for the multitude of photocopying they did throughout the meeting.

1 Site Reports

1.1 DAPNIA

DAPNIA is part of the Saclay institute and runs a mixture of UNIX workstations as well as a Cray and an IBM mainframe. This variety was a result partly of history and partly of free user choice. Most of the UNIX workstations were SUNs but many other architectures were represented; there were also some 20-30 X terminals. Use of X terminals was not without problems, especially when trying to provide good keyboard mappings for access to the IBM VM system.

There was a mixture of shells and GUIs (graphical user interfaces), including both OpenLook and Motif on SUNs. They were currently testing the CERN flavour of NQS.

1.2 University of Braunschweig

Architectures represented at the University ranged from an IBM 3090 to UNIX workstations. The latter first appeared some years ago and have increased rapidly in number recently such that there were several hundred now on the campus, mainly HP 9000 series 400 and 700 but also more than 100 SUNs and a few others.

Backup was still an open question: should they move to IBM's WDSF/DFDSM or not? They require an HP-UX client which the WDSF package does not have unless they use the CERN-produced code. NIS was in wide use.

They would like to add some large central servers, for example for file and compute servers.

1.3 IN2P3

UNIX has only recently arrived in IN2P3, in particular in the BASTA project and among the Engineering community where there were NeXT and RS/6000 stations. BASTA consisted of some 14 HP Series 9000/700 Model 730s, soon to be upgraded to Model 735s, and 3 IBM RS/6000 Model 550s, with more IBM nodes coming along with support for more I/O and for direct tape staging.

They have defined an IN2P3 unit equivalent to an IBM 3090 model 600S CPU. Using these units, their IBM mainframe produced some 7K CPU units per month while BASTA provided 13K, increasing to 25K after the upgrade. They would like to add more I/O capacity to BASTA as well as support for tape staging. They centralise UID/GID allocation via their IBM system and batch job limits are applied to BASTA use.

Elsewhere in IN2P3 there were laboratories with some SUNs and some DECstations, used for a variety of purposes including CAD. The labs have requested support from the Centre.

New projects being discussed included an interactive session server and a distributed successor to VMSETUP. Current concerns included the future of 3480 products and the usefulness of UNITREE or another file server package for storing home directories. They would like to promote the use of WDSF for file backup and to be able to provide reference platforms for software development and software distribution.

1.4 FNAL

To place it in context, FNAL has some 2000 employees and an even larger number of visitors connected via a large number of high speed national and international links as well as large onsite Ethernet and FDDI LANs. There were two major workgroup clusters consisting of the D0 and CDF experiments as well as many more smaller groups just getting started with UNIX. The main supported workstations were IBM RS/6000 and SGI but there were also some SUN and DECstation services offered.

FNAL currently had three separate batch projects - respectively one using processor farms (the so-called "factory batch", CLUBS and FNALU. The batch farm split jobs so as to use multiple CPUs for a job whereas CLUBS devoted single CPUs to individual jobs. The farms offered some 10K MIPS of CPU power (SGI and IBM combined).

CLUBS was intended as a replacement for the Amdahl mainframe and was being built up towards 1000 VUPS (VAX units); it should have good I/O services. They were going to look at a UNITREE-based system to replace the file backup/recovery functions of the Amdahl.

FNALU was coming online at the present time and was intended for workgroup support, interactive work and small scale UNIX batch (e.g. for development work). Currently only IBM RS/6000 equipment was installed because it was AFS-based and the AFS port to SGI was not yet ready.

Judy Nicholls also described briefly the FNAL Product Support scheme, principally UPD and UPS and also FUE, the FNAL UNIX Environment where the current effort was to add support for Bourne and Korn shells.

1.5   University of Dortmund

The main UNIX cluster installed here consisted of IBM RS/6000 systems linked by a serial optical connection. There were many X terminals also. The cluster had many roles from batch servers to software repositories and home directory servers. Since the previous HEPiX meeting, they had been able to perform speed tests in the cluster using the serial link and saw up to 5MBps with low CPU overheads on a single connection and up to 10MBps with 5 connections, but the CPU overheads climbed to 90%. One continuing problem was frequent crashes and these were blamed on automount so this has been discouraged and the use of static permanent NFS mounts with only occasional automounts has raised stability.

CERN's SHIFT tape handling software had been installed but its use was optional and TMS was not used; also multi-file support for Exabytes was missing. NQS was used but they had added local fixes for flow control and they were looking at adopting BQS instead. WWW was used but mainly for internal accesses.

They found X terminal setup a major problem; they had a large range of X devices including Apollo and PCs as well as real X terminals. They made some use of parts of DESY's X setup.

1.6   RAL

Since the last HEPiX meeting, RAL had updated its Cray system and was testing OSF/1 on Alpha workstations. User pressure would determine the OpenVMS to OSF/1 split on these systems but a small cluster was currently being purchased. An FDDI backbone was being installed and SuperJANET was planned to offer FDDI access speeds to remote UK sites in the near future. Some 40GB of disc space had been added for NFS access and a central mail-hub installed.

NQS was installed but mostly only for local use apart from a small 3 node HP cluster and WAN access to a remote Cray. An ASIS clone was installed and a scheme for ftp updating from the CERN master system was under test.

RAL's Virtual Tape Protocol was finding many uses, for example to offer VMS and UNIX access to tapes on the IBM and staging via RS/6000 discs. It also offers both physics data and file backup access to such clients.

SUE, RAL's name for their UNIX environment, was attempting to follow the current HEPiX study group as far as possible, trying for manufacturer independence and a standard naming convention for home file directories. It would be a package of scripts but the method of installation was unresolved, perhaps ASIS could be used.

1.7   NIKHEF

WCW (Scientific Centre Watergraafsmeer, of which NIKHEF is a constituent) was becoming a major Internet hub with links to EBONE, EUNET and a number of Dutch networks. The LAN was still based on Ethernet and Apollo Token Ring but there were plans to look at a fibre technology, perhaps ATM. Many workstation vendors were represented on site but mainly HP (both Apollo and Series 9000/700) and SUN.

UIDs were managed centrally and there was global NFS access and standard file structure naming to try to unify the environments.

Among the problems cited were too few support staff for the many different platforms, hard-mounted NFS file systems, user confusion as a result of too much personal customisation of environments and too much freely-available public domain software.

1.8   CERN

Since the previous meeting, the UNIX Workstation Support section had added preliminary support for Silicon Graphics workstations and built up further support for IBM RS/6000. Its major new activity however involved support for X terminals which were becoming very popular in CERN. A first issue of an X terminal administators guide had been published and support staff trained in installing the hardware and software for X terminals.

Development activities had included evaluation of Kerberos (still not mature enough to be put into general production) and AFS. AFS was to be covered in another talk at this meeting but it had expanded greatly in the past 6 months and was expected to continue to do so.

The problems of support remained as before and resembled closely those reported by NIKHEF above. One new one however was the change from SunOS 4 (also known as Solaris 1) to SunOS 5 (Solaris 2); many SUN and non-SUN applications did not yet work with the new software and Solaris 2 early releases were found to be buggy. This confusing situation is not expected to be resolved quickly.

1.9 DESY

The Apollo Token Ring network at DESY had now been frozen and most Apollo systems were only used as X terminals. On the other hand, the number of HP Series 9000/700 systems was rising; these gave some network problems (load) and there were plans to move to FDDI. Also the SGI farm would be upgraded to use the MIPS 4000 chips.

Other hardware changes: an Ampex robot will go into production shortly on the SGI farm for a long evaluation by the H1 experiment; around 80 printers were now accessible from all workstation platforms; and the number of X terminals was rising steadily.

Associated with the X terminal popularity much work had been invested in a standard X environment as well as on the common UNIX environment for HEP users; this work would be covered in the afternoon session in more detail. Among the spin-offs were a common NIS (YP) user registry, common login scripts, emacs as a supported editor everywhere and standard recommended keymappings for emacs.

DESY were looking at BACKUP site-wide and the principal candidates were Legato Networker and DFDSM from IBM.

Current problems included HP overloading (too many X terminal users and too much cluster traffic); the doubling of the X terminal population in the past 6 months; poor UNIX documentation; and user migration off the mainframe.

1.10   College de France and HEPiX-F

There had been 4 meetings of HEPiX-France thus far, most participants coming from IN2P3 sites but DAPNIA was expected to join shortly. The next planned meeting was at the end of May. There was a working group to study the questions of unique UIDs/GIDs within French HEP, starting from a collection of all relevent password and group files; the plan was to produce simple scripts to change UIDs or GIDs where necessary. HEPiX-F decided to join OSF and LAL Orsay subsequently did so on their behalf. They have established some WWW servers but were unsure about future support for the product.

Turning to CDF, the network there included a central NFS file base on a VAX as well as network-wide access to central printing. They have established a call-back system for access from home and were looking at PC-Remote to allow PCs to be used for this.

Several makes of X terminal were present on site and Xdm was used for display management; mwm was used for window management but they were unsure if a local window manager was indeed better than a remote one; it required a host daemon and was felt to be less reliable. When asked about configuring X terminals, the speaker replied that in his opinion host memory was more important than host CPU and suggested a minimum figure of 4MB per X terminal.

1.11 HEPiX US

Judy Nicholls reported on the meeting in March at CEBAF of the North American chapter of HEPiX. It had been a good meeting with lots of interaction. At the next meeting, tentatively scheduled for the autumn, it was hoped to try to encourage more user attendance as opposed to mainly system administration staff. There had been a request to cover more about futures and to feature some specific topics. An unanswered question concerned the "Tools Database" group, established in name at least at the HEPiX meeting last September and described in the HEPiX FAQs. What was it, what was its relationship if any to FREEHEP and who was doing what for it? A.Silverman replied that it required people and time to devote to it but he would try to re-activate it.

Another hot topic was the FORTRAN underscore and the confusion of calling sequences between C and FORTRAN routines. G.Folger presented some overheads on how CERN Program Library resolved the problem; the CERN Program Library includes tools to create the correct header files to permit C calls to FORTRAN sources. For calls in the reverse direction, it was proposed to use C macros. Regarding the underscore, many UNIX F77 compilers inserted underscores in subroutine names while others offered this as an option and so the recommendation was to always take this option where possible. CERN was approaching certain vendors to go along with this policy but VMS was a known problem, neither underscore by default nor as an option.

However, it remained the case that given a C object library, FORTRAN calls may be a real problem and we needed to persuade software providers to design their code properly, using stub calls and appropriate macros. The message should be passed to Les Cottrell of SLAC who raised the issue at the US meeting to contact the CERN Program Library team directly to explain the current problem.

2 Common UNIX Environment

2.1 Introduction

Alan Silverman introduced the topic of the afternoon session by giving some background on why the Common UNIX Environment Working Group had been established, what its activities had been thus far and what was being made available for discussion and initial testing.

The goals were to offer a flexible, tailorable working environment to HEP users which would be present at all sites as an option (perhaps at some sites as the default) environment for new users; it must be independent of users choice of shell and machine architecture. It was not intended that the proposed environment become compulsory usage, nor would it dictate which shell should be used.

The audience were reminded that, especially at the previous meeting in CERN in September 1992, there had been many calls within HEPiX for some standardisation in environments, especially catering for new HEP users migrating away from mainframes to the UNIX world, as well as some pleas from system administrators in smaller sites making a similar transition and asking for advice in how to setup UNIX for their users.

The Working Group had therefore produced and published some implementation guidelines and people at DESY had produced working scripts which were currently being tested locally and which would be offered for general testing within a very short time after the meeting.

2.2   Shells

John Gordon summarised some shell features which had been taken into consideration. It was noted that modern shells tended to be built on top of more traditional ones, incorporating their best features. However, many of the most recent shells existed only in public domain versions and therefore were not universally present on UNIX systems throughout HEP sites. Even shells built upon POSIX compliance could look different from each other.

Different shells have different startup sequences and scripts intended to be shell-independent must take account of these differences. And once again there were significant differences between scripts running under BSD and System V. Finally, we must remember that the sequence of calling startup scripts was different depending on whether we were starting a login session, a remote session and so on.

The solution being proposed was therefore to provide template scripts which simply source common scripts with sufficient checks to avoid re-calling a startup script which has already been executed to establish the environment. This had the extra benefit of restricting all the suggested environment setup to a small number of centralised scripts so that the users could feel free to edit their local startup scripts, apart from one line used to source the central script; this would thus permit individual tailoring for those who require it and also ease the situation where the common environment setup script had to be updated from time to time.

2.3   Variables and Scripts

Wolfgang Friebel then presented the current state of the proposed scripts. There were two sets of scripts, one for C flavour shells and the second for those derived from the Bourne shell. The variables to be defined were split into three classes: some necessary such as PATH, TERM and so on; some were desirable such as EDITOR, PRINTER, PS1; and the last set were suggested so as to produce the level of desired standardisation and include such variables as OS, SYSTYPE, SHELL, etc. A special variable HEP_ENV must be declared which defines whether the common environment has already been set or not; this is the variable used to avoid re-sourcing the common script.

There was the suggestion to add CDPATH to the list of useful variables but some people preferred to keep this as a local variable. It would be nice to declare keyboard type and display type but it was thought very difficult to be able to automatically set these correctly in all cases. Although it would be preferable to avoid asking the user questions at login time, it might be necessary if these were to be defined in every case; an alternative was to give the user a message to say that some variables could not be reliably set and suggest a command structure to set them by hand.

There was also a discussion about whether or not to define aliases and it had been decided to declare a few, mainly as examples for new users. Similarly, a default prompt string should be set and

command line editing, the history mechanism and filename completion enabled for those shells which supported these features. It was emphasised that the prompt string would be provided mainly as an example of what could be done.

The common scripts being produced were a compromise between modularity, speed of execution, readability and size. When ready, they would be placed on an ftp server at DESY and publicised. Sites would be encouraged to copy them and test them and feed back comments to the authors. It was agreed that we could not hope to merge existing users' startup files into the proposal and the best to hope for was that such users wishing to participate in this common environment proposal should source the common files first and then continue with their own setup.

2.4   X11

Thomas Finnern described some of the work done at DESY to support X11, especially with regard to use on X terminals. The R2 group there has about 300 X terminals and workstations using X. Most of their X terminal users connected to either Apollo or HP Series 9000/700 servers but the use of SGI was expected gradually to replace the Apollos.

DESY had decided that freely-accessable X terminals were insecure and had decided to support the MIT "magic cookie" scheme. Unfortunately this was not available on VMS. A number of administration tools were provided including SNMP-based utilities for status query and control.

From the point of view of user support, procedures to customise an X terminal had to be simple to use; good documentation was essential. An Xdm login panel was provided based on a Motif interface and a standard virtual keyboard; keyboard definitions were provided for this along with virtual font names.

Among the user tools provided were those to establish an Xsession with a standard user environment; a method to map the DISPLAY variable correctly even when doing remote connections; xctrl for control tasks including locking of idle terminals; x11init to debug X resources and to be able to re-initialise corrupted terminal settings; xrsh using the MIT "magic cookie"; and various X terminal configurations for different terminal emulations.

In answer to a question as to why DESY users were choosing X terminals rather than workstations, the speaker cited the relative costs of the devices, the fact that workstations involved the user in some system administration tasks and the cost of software licensing. He did not think there was much noticable network traffic difference between the use of local or remote window managers in many simple cases.

2.5   Keyboard Mappings

As mentioned in the previous talk, DESY had also invested effort into trying to standardise keyboard mappings and this was described by Axel Koehler. The goal was to offer generic terminal handling by providing a database to describe the capabilities of the different keyboards and then supplying

subroutine calls to query this database. It was noted that alphanumeric terminals were radically different from graphic ones in that they mapped a key to a code while graphic terminals return a pointer to a symbol in a table. The non-alphanumberic sections of the keyboards were split into four sections - arrow keys, function keys, application keys and special function keys; keyboards fell into two varieties - PC-like or VTxxx-like.

The guiding principle behind the mapping chosen was "what you see is what you get"; thus where possible, a key returns the corresponding character or behaviour engraved on the key. The exception was the numeric keypad which was used in application mode, that is position-dependent; more details were given on the foils.

Within gnu emacs, key use was dependent on whether the window was performing terminal emulation or native X. Emacs has been patched to provide virtual keymap functions when running in native X mode. A full set of mappings for emacs was now in use throughout DESY. Other key mappings were offered for some browser utilities and for shell command line editing. Experience thus far had been generally positive, new users particularly finding them easy to use and helpful for "getting into" UNIX.

In the discussion of the question of why certain choices had been made the fear was expressed that other labs might make different choices; would it be possible to merge all these? It was pointed out that someone at CERN (L.Cons) had done some work with respect to EDT emulation; he should be encouraged to publicise his work. It was further pointed out that the next version of emacs would probably involve major changes. Finally, the need for interworking between VM, VMS and UNIX must be kept in mind.

2.6   General Discussion

It was agreed that users should be allowed some choice, perhaps somewhat constrained by local system management and that the proposed common environment must not restrict this choice. The question of merging in existing users' working environments had no simple answer. Not directly related but the suggestion was made to try to organise emacs tutorials at future HEPiX meetings for those interested; and other such topics could alse be considered such as PERL, TK/TCL, etc.

3   ASIS

Philippe Defert, the author of ASIS presented the main purpose of the project, to provide a central store of public domain software and to make this available for copying to distributed servers or for direct access by end-users. The project was now running in production mode in CERN with a clone in operation in RAL. At CERN there were already many hundreds of users, some from outside CERN in order to access the latest release of CERNLIB. Typical access methods were anonymous ftp for copy and NFS and AFS for access directly off ASIS discs.

Plans for the future centred around extending access via AFS rather than NFS although some continued NFS and ftp access can be expected to be required for some time. Within a short time, the

master copies of ASIS programs would be stored on an AFS server with the copies for use by NFS being updated from the AFS copies. Later, when AFS has proved to be stable in production, it should be possible to drop the NFS copy but access the files directly from the AFS server using the NFS to AFS converter for those users still requiring to use NFS.

It had been found that there was an urgent need for more people to volunteer to maintain particular products so that the most up-to-date versions were always available. To assist this, some effort will be expended in providing tools and documentation for product maintainers. Another requirement was for access to computers of each supported architecture to compile and/or test new releases of software for storage on ASIS. Lastly, it may be time to rethink the ASIS strategy of basing product storage on specific UIDs; this would imply more sophisticated ASIS management tools and a more rational directory structure could be implemented.

Gunter Folger then explained how to access the latest CERNLIB releases via ASIS. Users are required to complete a registration form found in the anonymous ftp account of ASIS and should then read the README files in the relevent directories before copying the libraries. In the future, CERNLIB releases would follow ASIS in the trend towards AFS use although the question of how to pre-register AFS users was still open. Also being worked on were plans to offer access to CERNLIB documentation and READMEs via WWW.

4   HEPiX Support at DESY

Thomas Finnern presented some of the work done at DESY which had been developed in relation to the effort to prepare a standard HEPiX environment, as covered the previous day. The requirement was to support clusters as well as individual workstations. A procedure called SALAD had been built which could move files around as necessary, performing suitable checks; this could then be used for maintaining certain system files and layered products.

One thing such a scheme must take care of was different system architectures and thus the need to copy architecture-specific binaries. For each product or set of products, there existed a description file which listed all required files and special actions to perform on them during or after the copy, including executing scripts in some cases.

SALAD was being made available for public use and its development will be traced in the HEPiX news group. Already it was in normal use at DESY, although so far only for smallish products.

5   Password Checking at CERN

This was presented by Alan Silverman although the actual work was authored by Lionel Cons who was unable to be present in Amsterdam. In UNIX security, weak passwords were generally acknowledged to be the largest single problem. A check on some 72 UNIX machines at CERN had found the correct password in over 27% of the 4736 accounts, including 78 accounts without passwords. A service had been started using the CRACK public domain password checker to analyse password files of UNIX

systems and check for weaknesses. Rule files had been built to perform many checks (274 rules) on a number of dictionaries (140K strings). The service was only recently opened but was already very popular and some major UNIX systems had used it to "clean up" their account registry.

Continuing work involved devising good rules for initial settings of passwords, putting these rules into the passwd program and perhaps feeding such requirements back to UNIX vendors; and also adapting the PERL COPS module for general security checking of UNIX systems.

6 FNAL Developments

Judy Nicholls presented various development activities currently taking place at Fermilab.

6.1 CLUBS

CLUBS was a project to provide batch services on UNIX systems. Currently it was using FNAL's Amdahl processor to access tapes but there was a move away from this towards staging via RS/6000 workstations. They were hoping to use UNITREE on the RS/6000 as a possible access mechanism for data in their STK silos. The batch system used in CLUBS was based on Condor.

6.2 FNALU

FNALU will attempt to use AFS in a batch environment. AFS had been chosen as the preferred file system partly for performance reasons and partly to avoid complicated NFS cross-mounts. It also offered improved management and security features on the server and volume replication. However, both users and system administrators would need to become familiar with AFS. Another major problem to overcome was the limited lifetime of AFS tokens when used within long batch jobs.

6.3 UPS

After reviewing the design goals of their UNIX Product Services (rules for product file storage, uniform user access to products and documentation, easy distribution of products etc), Judy described some of the latest developments. The UPD system was used to copy product kits from the central repository to local disc but users needed to pre-register to use UPD. Tools were provided to install products and setup a user's environment to use them. UPP was developed to provide an automatic update interface for those who wanted that feature or a notify option for those system admins who preferred to update at times fixed by themselves.

6.4 LUE

This product, Local User Environment, was developed to collect information about remote systems on demand. Originally developed for UNIX systems, it had recently been extended for VMS use. It was effectively a mail alias for the execution of a shell script on a remote system which then returns information to the requestor. Local system admins have some options to decide what information may and may not be returned.

7 AFS

Rainer Tobbicke reported on the work being done in CERN to establish an AFS service. The most important features of AFS were presented, namely that it was a distributed file service with local disc caching of files; it used Kerberos user authentication; and it had a global filename space. As of that moment, CERN had installed two AFS servers with a total of 7GB of disc space devoted to AFS and 30-40 clients on a wide variety of workstation architectures.

As had been mentioned earlier, AFS was being used in the ASIS project for accessing the repository of public domain utilities. Currently, the AFS files were copies of the ASIS master but this would shortly be reversed and the AFS files would become the master copies. The full CERNLIB package was available via AFS in this way.

Some 35 AFS accounts had been created for file storage, mostly for home directories and such use was steadily increasing. Also a few projects (FATMEN for example) were moving to AFS. A CERN AFS User Guide was in preparation.

The AFS servers had a working backup scheme using an Exabyte Stacker but a special AFS backup utility was needed in order to be able to restore the full AFS file protection information (the Access Control List). Volume cloning was used in the backup scheme and this feature could also be used to improve performance and reliability for read-only files.

AFS uses Kerberos authentication and the supplier (Transarc Corp) had supplied some Kerberised utilities (login, ftpd, remsh for example). Rainer had added others to handle Kerberos tokens; these included Xdm, rlogin, xlock. This last one was most useful in prolonging AFS tokens for individual workstation use.

Transarc had defined a recommended directory structure which offered easy global file access between AFS sites. This had one major drawback however; a user starting a grep or find from the top level directory risked having to wait literally days for a reply! Using a special AFS notation (@sys) as part of the path name one could easily hide architectural dependencies when referring to a binary executable file.

Performance was broadly similar to NFS when a file was first accessed but subsequent accesses showed very large improvements, running at basically local disc speeds. Support from Transarc was good and stability was excellent. It was known that moving to DFS would imply changes in naming structure and perhaps in other areas and CERN were watching this closely.