Experiences supporting Linux at DESY Hamburg

Dirk Wetter

DESY Hamburg, Computer Center, Notkestr. 85, D-22607 Hamburg, GERMANY

Abstract

In the commercial market Linux running on commodity PC hardware is reaching critical mass. In High Energy Physics the Linux operating system is already established. There it proved in the last few years that it is well suited for desktop computing, for batch farms and for so called workgroup servers. Within some of these computing models it competes with the old established UNIX operating systems and also with Windows NT.

But compared to the already supported UNIX flavors at DESY it introduces more than the difficulty of a new UNIX operating system: The installation basis especially for decentralized desktop systems is larger, the technical development of this young operating system and the PC hardware is incredible fast. Moreover due to the widespread use of Linux special attention has to be focussed on a secure setup of a Linux PC.

Starting with full central Linux workgroup server and farm support, central computing at DESY developed an installation support for the personal use on the desktop as well. The challenge was and still is to satisfy the support requests with a minimum effort of human resources so that the low cost for the hardware should not be compensated by high cost for the support. This could be achieved using a hands-off network-based installation mechanism which takes less than one hour. The resulting system has the same functionality as other supported UNIX flavors at DESY.

1. Introduction

The motivation to establish another operating system in a High Energy Physics computing environment are different from the collaborations point of view and from the computer centers point of view: both of them want to have a stable and reliable UNIX operating system, Linux as of now (kernel 2.0.35) basically fulfills this point well. The Linux operating system running on Intel based PCs has a good performance and the hardware is cheap. But for the computer center which has to provide a service for an operating system and for the hardware the total cost of ownership (TCO) [1] is the more important issue. The TCO includes

apart from hardware costs as well costs for preparing, maintaining and performing the installation, for the administration as well as technical and user support. Moreover the application software has to be provided and distributed.

Especially for Linux desktops there are more frequently used arguments from the collaborations: e.g. having a more homogeneous UNIX environment accessing workgroup servers or e.g. SMP batch or disk servers. There’s also the effect of having a ’personal’ computer on the desk.

Fig.1.: DESY computing model

At DESY existing different computing models: SMP compute or SMP disk servers, workgroup servers accessing these servers, batch or farm nodes and the individuals’ terminal accessing these kind of computers: the desktop device. Linux seems to be a multi purpose operating system: apart from the SMP disk servers it’s well suited for all computing models. Whereas e.g. Windows NT is rather used for desktop computing (at DESY) and e.g. Solaris/Sparc is the classic workgroup server platform.

2. The concept

2.1. Workgroup servers

A workgroup server at DESY Hamburg is defined as a server which offers services for a workgroup using this computer. These are up to ~30 people. UNIX services like mail, news or web are hosted by special central servers where no interactive work is allowed.

The distributed filesystem plays a keyrole at DESY for the workgroup server environment. Currently AFS is used for distributing the home directories (integrated login with a token) and the application software (e.g. netscape, emacs, TeX, paw, gcc). This means that only the basic system system software is stored locally on disk. From the administrators point of view this makes their tasks of providing a reliable and homogeneous service easier. From the users point of view the distributed directory trees via AFS are a part of the "DESY look and feel". Another part is the "DESY environment" (HEP-/login- and X11-environment, setup of mail, printing). Both of them guarantee the users the same homogeneous working environment independent of their UNIX flavor.

Workgroup servers do not have a local display or a keyboard. They

are centrally located either in the computer center or in dedicated rooms. A high bandwidth network (FDDI, fast ethernet) is needed for these machines, basically they have NFS access to data disks via automounter. To guarantee a certain level of service the workgroup servers are monitored.

Linux workgroup server support means that the expected number of additional computers are low, probably a few 10 Linux PCs will replace existing workgroup servers from other UNIX brands. The functionality and the setup of these machines is nearly the same. They can be easily administrated with existing tools and concepts. Since spring 1998 the DESY computer center has supported Linux workgroup servers. The level of support is the same as for the established UNIX flavors: Central computing provides installation support, the administration is shared with the local group administrator, the application software is available and the users can contact the User Consulting Office in the case of questions or problems. 17 workgroup servers are currently installed, additional 18 machines are batch nodes but have a workgroup server like setup.

Sometimes batch machines or farms nodes have special hardware or software requirements. E.g. 20 non-centrally supported farm computers don't run AFS and YP. Or the cluster has to have brand new kernel features which are only available in the so-called hacker kernels (currently the 2.1 series). These individual solutions need some human resources which the computer center cannot provide, Central Computing supports only machines with a workgroup server setup.

2.2. Desktops

Desktop computing in general means that there is a large number of geographically distributed systems. Therefore a good concept is needed in order to support any kind of desktops centrally.

At DESY Hamburg there are currently about 1200 X terminals. Because of their technical simplicity the central support is a relatively easy task. But also because of this technical simplicity the possibilities and the usabilities are restricted.

Supported at DESY are also PCs running Windows NT. The functionality of the X terminal is replaced by a commercial X server (Exceed) and one can run locally web browsers or office applications which disburdens the UNIX servers. AFS could be optionally installed on the NT PCs.

Interesting is the NT administration model at DESY: It's being distinguished between green, yellow and red NT PCs. Green means that the user doesn't have any privileges on his machine, red means that he can do whatever he likes with his PC, it is used only the installation method [2] and it's in the DESY NT domain.

For Linux however there were (and still are) in the past lots of self installed systems on the individuals desk. Not only a few of them were badly installed from the security point of view so that breakins took place and endangered the operation of other computers. If one compares

[4] the number of exploits for different UNIX platforms, Linux scores as the "best". Thus for a large number of decentralized Linux desktop systems special attention has to be focussed on a secure setup of the PC. In order to guarantee the secure setup after the central installation there have to be rules for the operation.

At DESY support for Linux desktop systems is very often requested. An estimation yielded some hundred PCs. Thus a well scalable solution (from the organizational as well as technical point of view) is needed. Hardware standards make the support much easier. At DESY a group is in charge of defining in certain intervals the so called "DESY standard PC" [3]. This PC has a certain network card, a certain graphic card, a certain motherboard and so on. The functionality is guaranteed for all supported platforms.

The setup of the Linux desktop system is based on the Linux workgroup server. The system has as well the DESY look & feel (AFS, DESY environment). The basic differences are the support of local in- and output devices, the absence of automounts to data disks and a more restricted access to this computer. However in order to keep the load for the computer center low, for hundreds of PCs the level of central support has to be different from the workgroup server support. Thus only an easy to use network installation is being provided. The installation and administration has to be performed by the collaborations who have prior to the installation procedure nominate a group administrator. The rules for operation of these desktop system are comparable to those for the green Windows NT systems: e.g. users don’t have root privileges, root and user access has to be granted for the computer center all the time. This means also that dualboot systems are not supported. Currently at DESY Hamburg there are more than 40 test installations. Production will start within 1998.

2.3. Prerequisites

The PC hardware undergoes frequent changes. For the Linux developers it’s hard to keep track with this rapid development especially for brand new network-, SCSI- or graphics cards. For a service provider like a computer center it is even more difficult to keep track of the kernel or driver development because the driver or the kernel release has to be tested, the central installation repository has to be modified in order to distribute or at least provide the recent release. The procedure of applying, testing and providing causes load which in other cases is undertaken by the operating system vendors. For other UNIX flavors there’s often not a big choice from which vendor the hardware should be bought. For PC hardware however the choice is made very often for the cheapest but not best offer. In some cases this lead to hardware which was simply badly assembled or e.g. after one year there were broken CPU fans or broken harddisks which turned out not be suited for long duration operation.

Linux is a young operating system. The kernel which is being used now (in fall 1998) is basically more than two years old, though a development with certainly some improvements took place from kernel 2.0.0 to the currently used 2.0.35 kernel. The future user kernel 2.2

however has a lot of basic innovations [5]. E.g. as of 2.1.30 the SMP code [6] was rewritten from scratch. The NFS code was revised [7], resulting in much better performance, the same for memory management, networking and PCI subsystem [5].

There are also changes going on this year for 3rd party software: Although for the Linux operating system a lot software is still not available or only available with a reduced functionality (batch facilities like Loadleveler or LSF, DFS client), some companies made or will make also their software available for Linux within this year. Some examples are Objectivity, Oracle, IBM (ADSM), Sun (JDK) or even Transarc, which recently announced to provide AFS 3.5 clients for the upcoming Linux kernel 2.2.

Up to now there was no support from Transarc for the Linux AFS client. A graduate MIT student was developer and maintainer [8], the AFS version was an outdated one (3.3a), there was no guarantee that once DESY upgraded their AFS servers to the upcoming 3.5 release the client will still work as it should. The Linux AFS client distributions were not complete, some more (for remote access: rsh, rlogin or inetd.afs) or less (package, dlog, scout) important binaries were missing. The AFS libraries were due to export restrictions not available outside the US, hence it was also not possible to compile the important binaries. Because in the DESY UNIX environment AFS plays a keyrole, the status was not satisfying.

3. The method

DESY uses currently the SuSE distribution. The history behind that is that DESY looked in fall 1997 for an easy to use network based installation mechanism, comparable to e.g. the Solaris jumpstart procedure. Later that year the RedHat 5.0 distribution was launched, which provided a similar mechanism. The drawback was that for production systems a glibc based distribution was not the best choice: The AFS client and mosts of the HEP software (e.g. CERNLIB) was not available in glibc format. Thus DESY decided to go for SuSE and ordered the development of the network based installation mechanism. Meanwhile it was reintegrated into the official SuSE release 5.2. A disadvantage of this installation server is, that it is a "monolithic" one: The client gets everything according to the settings on the server, for the client there is only a small degree of freedom for influencing the installation procedure. This means that the maintainer of the installation server (the DESY computer center) has to do most of the work.

The SuSE distribution uses as well RPM as a software deployment tool but the naming scheme for their RPM packages is (only) 8+3. A further difference from the administrators point of view is the scheme of booting up the system. The more comfortable SuSE scheme uses a global configuration file /etc/rc.config, the startup scripts and the links are below /sbin/init.d (comparable to Digital Unix). A clear disadvantage is the lack of Pluggable Authentication Module (PAM) based authentication service [9,10]. For Linux exists a PAM-AFS module [11], which allows to get (xdm, login, ftpd) or to renew (xlock) an AFS

token. Since the PAM software is free, it replaces at DESY the SuSE authentication mechanism. For local entries shadow password are used. The password information (but not the encrypted password itself) is distributed via YP at DESY Hamburg. SuSE uses the newest libc5 library, which guarantees a more reliable and faster NIS service e.g. compared to RedHat 4.2 which provides a special version of an outdated libc5 (5.3.12).

The Linux installation server has a crucial role: The setup of all Linux computers which will be or were installed is stored there as a profile within a class definition. The criterion for the class is the IP number of the client. With this class it's being distinguished between the different clusters and the model for which the computer will be used for. For every class one can specify e.g. the partitioning scheme for the harddisk(s), the software packages, the kernel and modules, the cluster specific setup (netgroups in /etc/passwd, predefined root passwords in /etc/shadow, autofs maps, YP client configuration) and so on.

For the Linux installation at DESY a floppy disk is provided. After booting the latest generic Linux kernel, a bootp client, included in the ramdisk on the floppy, sends its request. The central bootp server checks the MAC address, identifies the system and sends the NFS servers IP number and the directory which has to be mounted. After mounting this directory the harddisk is partioned and formatted according to the rules on the installation server. Hereafter the basic installation of selected RPMs from SuSE takes places. Newer software revisions make it sometimes neccesary to update the SuSE packages. The SuSE postinstallation provides an easy mechanism to install own packages. In this phase e.g. security fixes are being applied, the modul utilities for the MIT implementation of AFS are downgraded, an own compatibility package to execute libc6 binaries is installed. The class definition specifies in this step which kernel, including the modules, is installed and which cluster specific RPMs are to deploy. In order to boot the selected kernel and to continue afterwards the installation process, a hook is created in the startup scripts. After the reboot the hook invokes the DESY postinstallation where AFS, AFS- and PAM-aware binaries including the PAM-AFS module as well as ssh are installed (as RPM packages). With salad [13] the computer gets finally the DESY specific setup (HEP-/X11 environment, access to all printers, mail setup). After the hook has been removed the systems reboots and is ready to go . Depending on the speed of the CPU, network, harddisk, the number of packages and the size of the AFS cache, it takes about 20-60 minutes after the initial powercycling of the computer. The whole installation process is a 'hands-off' installation, the computer gets basically its new system without pressing any key.

4. Conclusion

Central Computing at DESY has succeeded in a short time to raise a central Linux support for both computing models, workgroup servers and desktop as well. In spite of the fast development of Linux and the PC hardware, the DESY computer center has faced up and until now won the

challenge to keep the total costs for supporting Linux running on cheap commodity hardware low.

This could be achieved using on one hand technical solutions, namely AFS and a fast network based hands-off installation, on the other hand a distributed support model which also should be scalable for hundreds of PCs.

5. References

[1] http://webopedia.internet.com/TERM/T/TCO.html

[2] http://desyntwww.desy.de/noframes/hepnt/NetInstall.htm

[3] http://deskpcsc.desy.de/dtgpc/spc.html

[4] http://www.insecure.org

[5] http://www.linuxtoday.org/stories/296.html and

http://www.uwsg.indiana.edu/hypermail/linux/kernel/9703.3/0235.html

[6] http://www.phy.duke.edu/brahma/smp-faq/

[7] see file Documentation/Changes in the Linux 2.1/2.2 kernel tree

[8] Linux Journal, 53 (1998) 76-77

[9] Ted T'so, USENIX 1997, Annual Technical Conference and USELINUX, submitted notes, (1997) 209-217

[10] http://www.dementia.org/~shadow/pam.html

[11] http://www.uni-hohenheim.de/~schaefer/linux/pam/index.html

[12] http://www.gnu.org/software/libc/

[13] http://www.desy.de/~finnern/cc/desy/salad.ps.Z