

High performance Linux clusters

-

design/configuration/maintenance

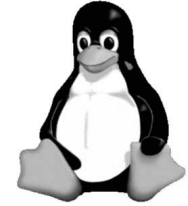
Werner Scholz

Theory Group

Roy Chantrell

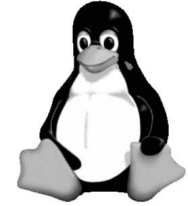
Jan. 3rd, 2002

Outline



- **Motivation**
- **Installation at TU Vienna**
- **Stability, reliability**
- **Performance**
- **Dual use: workstation and cluster node**
- **Data storage**
- **Queueing system, user interface**
- **Remote access**
- **Security**
- **Management**
- **Summary**

Motivation



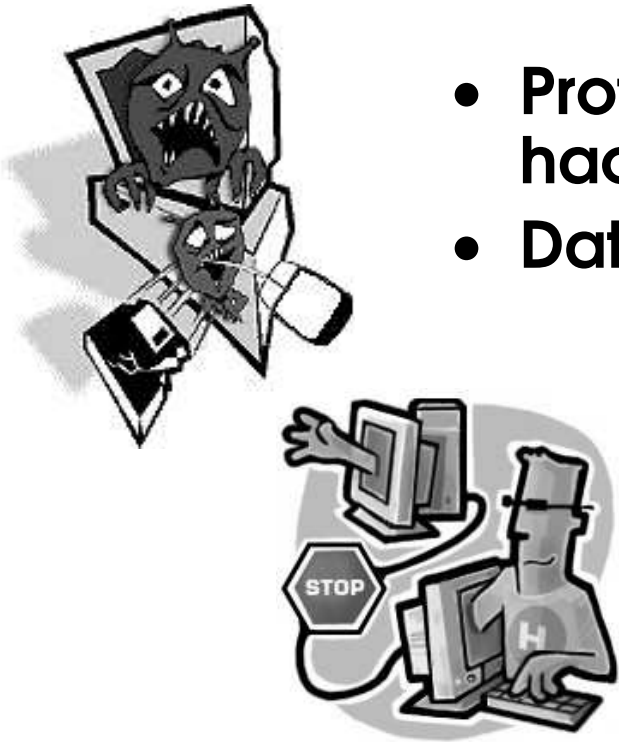
- Growing working group of Josef Fidler and Thomas Schrefl



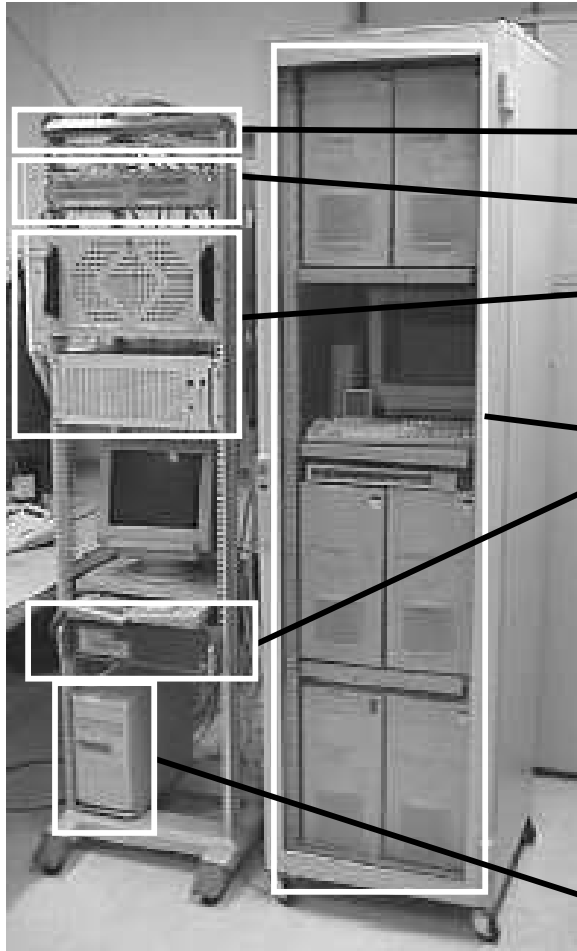
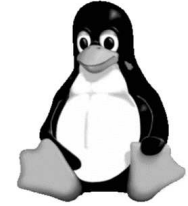
- Protection against hackers and viruses
- Data security



- High performance cluster for FE micromagnetic simulations

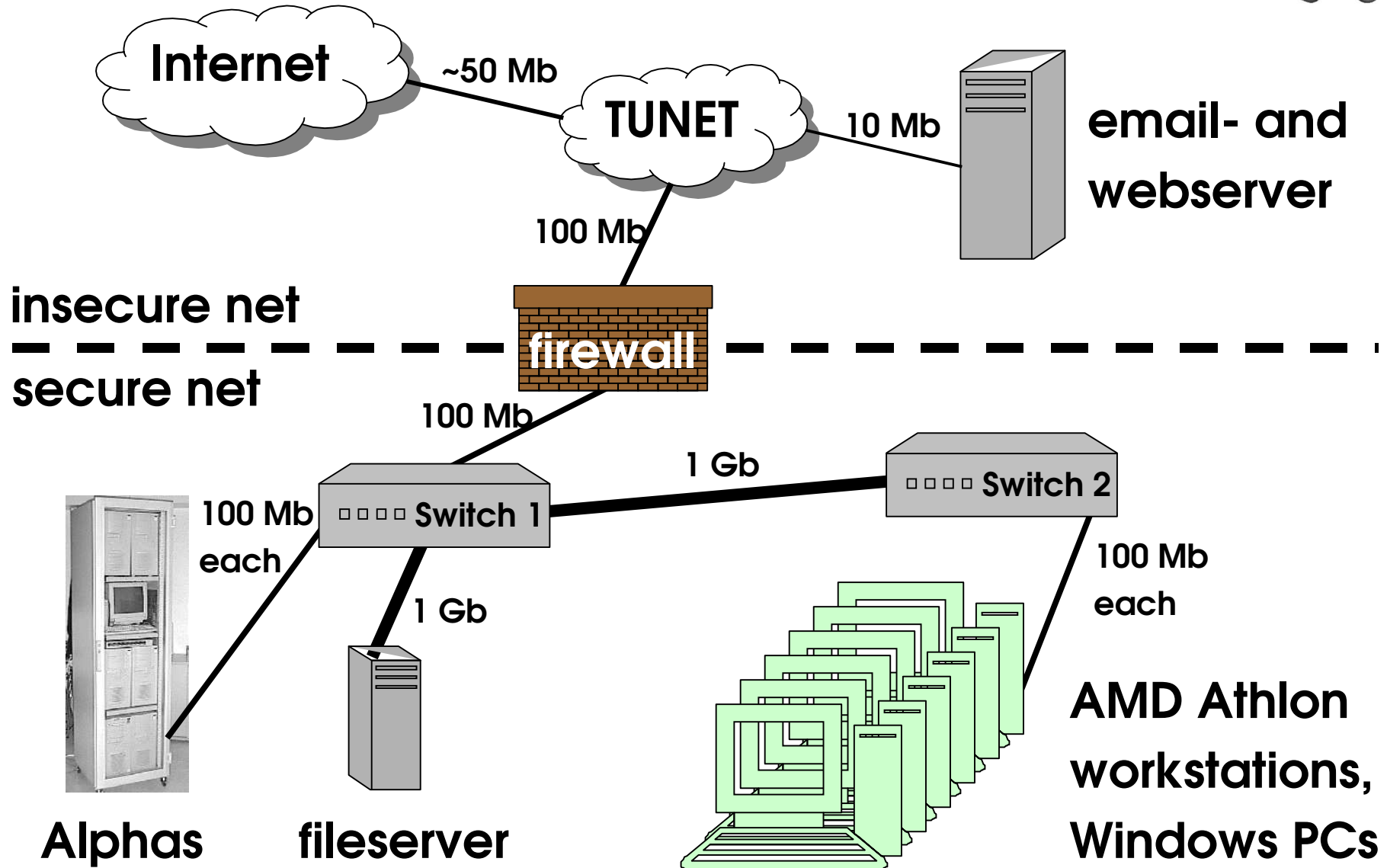
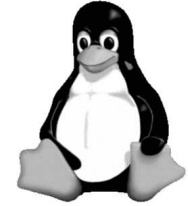


Installation at TU Vienna

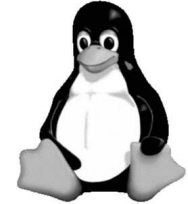


- **Secure network**
 - switch
 - firewall / Linux
 - fileserver / Linux
 - 7 AMD Athlons / Linux
 - 7 Alphas / Linux
 - 1 DEC Alpha / Tru64
 - Windows PCs, notebooks
- **Public access**
 - web- and email-server / Linux
 - <http://magnet.atp.tuwien.ac.at>

Network layout



Stability 1



- Network in operation since Jan. 2001

- Firewall

- intel Celeron @ 633 MHz
- debian Linux 2.2



problems until Jan. 2002:

defective harddisk (detected before failure)

- Fileserver

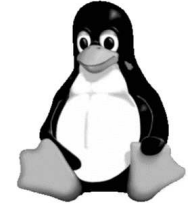
- 2x intel Pentium III @ 733 MHz
- SuSE Linux 7.0
- Linux Kernel 2.2.17 RAID SMP



problems until Jan. 2002:

recently 2 reboots for file system problems
~5 reboots for power down, installation of
additional harddisks

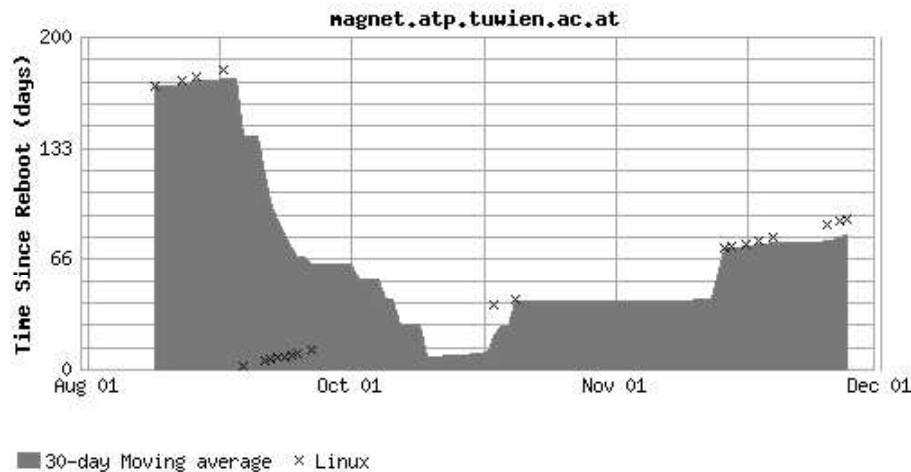
Stability 2



- email- and webserver
 - intel Pentium I @ 166 MHz
 - RedHat Linux 6.1
 - Linux Kernel 2.2.13
 - Apache mod_ssl 1.3.22
 - Sendmail 8.11.6
 - openssh 3.0.2p1
 - Kaspersky Anti-Virus



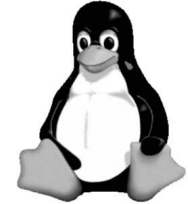
- Reboot on 09/15/01 after installation of additional harddisk
- Performance problems with secure Webmail interface (Neomail written in Perl)
- Recent security updates:
 - Apache
 - Sendmail
 - ssh



(c) Netcraft, www.netcraft.com

Working Group MAGNET
<http://magnet.atp.tuwien.ac.at/>

Performance

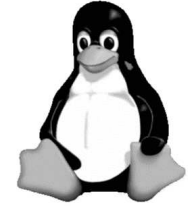


- Simple benchmark with a finite element micromagnetic simulation
- CPU
- Memory
avoid swapping
- Network bandwidth
Ethernet, Myrinet
- Requirements strongly
problem dependent
 - serial processes running parallel
 - parallel processes with little interprocess communication
 - parallel processes with heavy interprocess communication

processor	Speed (MHz)	time (s)
AMD Athlon	700	10899
AMD Athlon Thunderbird	900	10055
Intel Pentium III	500	16252
Intel Pentium III	733	12000
Alpha EV 5	333	8500
Alpha EV 56	533	8346
Alpha EV 67	667	2486



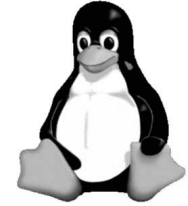
Dual use: workstation and cluster node



AMD Athlon workstations

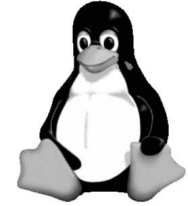
- Hardware
 - 900 MHz
 - 512 MB RAM
 - 20 GB harddisk
 - 2 GB swap space
 - 100 Mb/s network connection
- Software
 - RedHat 6.2, Kernel 2.2.17
 - Gnome desktop (or KDE, or...)
 - StarOffice 5.2
 - Netscape, Mozilla, Opera
 - Acrobat Reader
 - Maple, Mathematica
 - LaTeX, GhostScript
 - XMGR, Grace
 - xfig, xv, Gimp, PovRay
 - GNU compilers, tools, DDD
 - VMware (Windows 98)
- Performance
 - most of the time CPU is idle
 - lower priority for background processes assures fast response for interactive work
 - best use of available CPU time
- Problems
 - large simulations (better avoid swapping!)
 - bottleneck: not the network but the harddisks of the fileserver (NFS access during file compression, backups on the fileserver)

Software installation

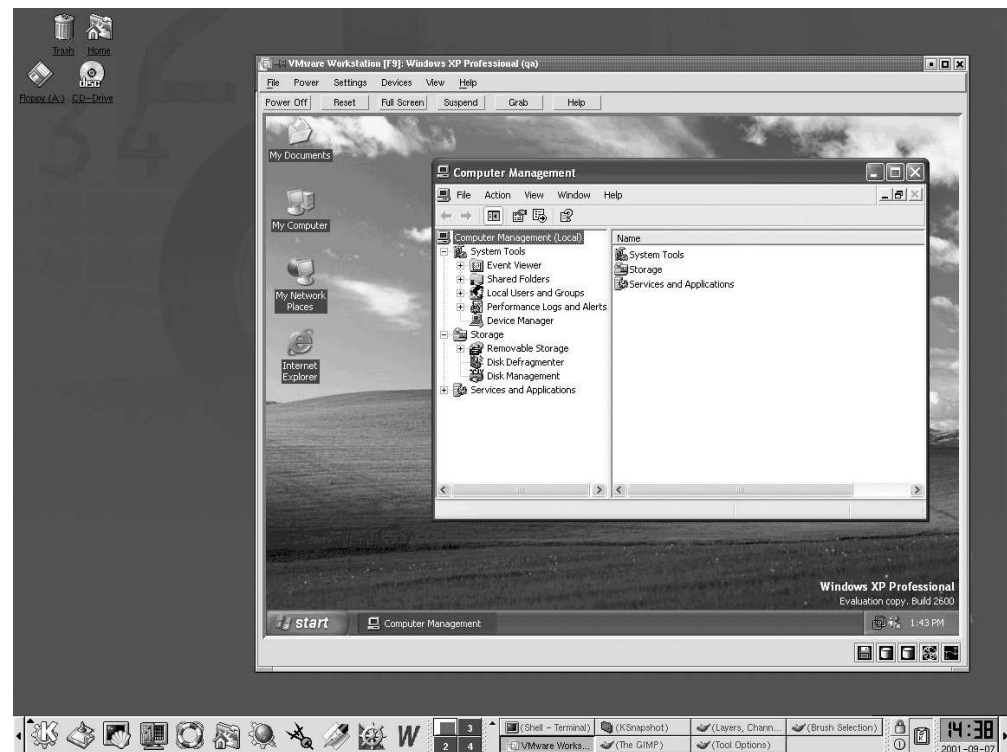


- **Standard RedHat 6.2 installation**
- **all additional software is installed on the fileserver and mounted by the clients**
- **greatly simplified software maintenance**
 - **single installation of every package**
 - **simple updating**
- **Simple adding of new clients**
- **full remote access for root to all clients:**
issue 1 command and execute it on n clients

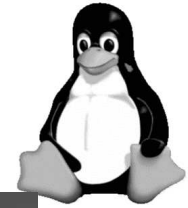
Windows and Linux



- When you have to open MS Office documents:
 - StarOffice (free also for commercial use)
 - rumors about a Linux port of MS Office
- Run Windows programs within Linux:
 - Wine:
implementation of the Windows 3.x and Win32 APIs
 - Virtual PC
 - VMware: boot a virtual machine

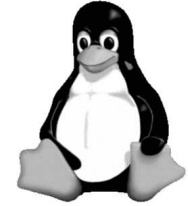


VMware



The screenshot displays a VMware Workstation interface running a Windows XP Professional virtual machine. The main window title is "VMware Workstation [F9]: Windows XP Professional (qa)". The desktop environment includes icons for "Trash", "Home", "Floppy (A:)", and "CD-Drive". A "Computer Management" window is open, showing a tree view of system tools and storage. The taskbar at the bottom of the VM window shows the Start button, "Computer Management", and system tray icons including the clock showing 1:43 PM. The bottom of the image shows a host OS taskbar with various application icons and a clock showing 14:38 on 2001-09-07.

Lotus Notes with Wine



The screenshot shows the Lotus Notes application window titled "Welcome - Lotus Notes". The interface includes a menu bar (File, Edit, View, Create, Actions, Text, Help), a toolbar, and a navigation pane on the left. The main content area is divided into several sections:

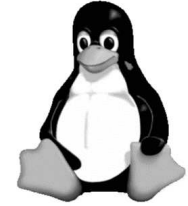
- Inbox:** A list of messages with columns for sender, date, and size. The date for all messages is 11/13/2000.
- Calendar:** A section for Friday, November 24, 2000, showing "No calendar entries for today".
- Search:** A search bar with a "Go Get It!" button and a "Search for" input field.
- Edit News:** A section for "November 24, 2000" with "Top News" and "Business/Reuters" links.
- Edit Stocks:** A section for stock market data, including a table of stock prices and a "Wingspan Bank.com" logo.

Symbol	Last	Change	% Change	Volume
DJIA	10,470.23	+70.91	+0.68%	N/A
NASDAQ	2,904.36	+149.04	+5.40%	N/A
IBM	99.94	+1.44	+1.45%	2,497,500
TRLY	15.66	+2.19	+15.96%	1,519,600

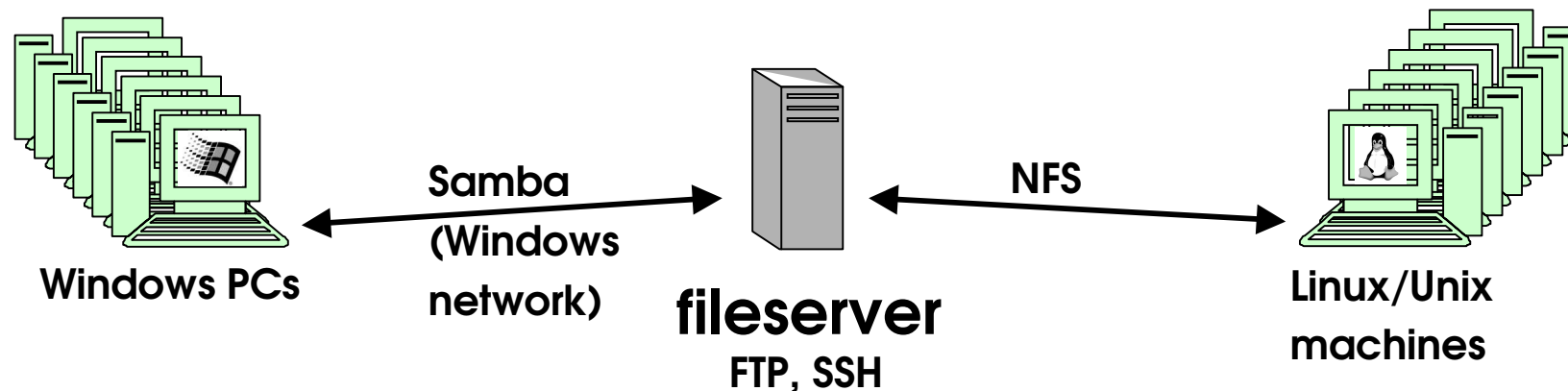
Updated: 6:56:52 PM EST 11/24/2000

Quotes delayed 20 minutes. Mutual Fund NAVs are as of latest market close.
Data from S&P Comstock.

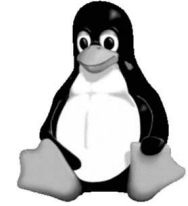
Data storage



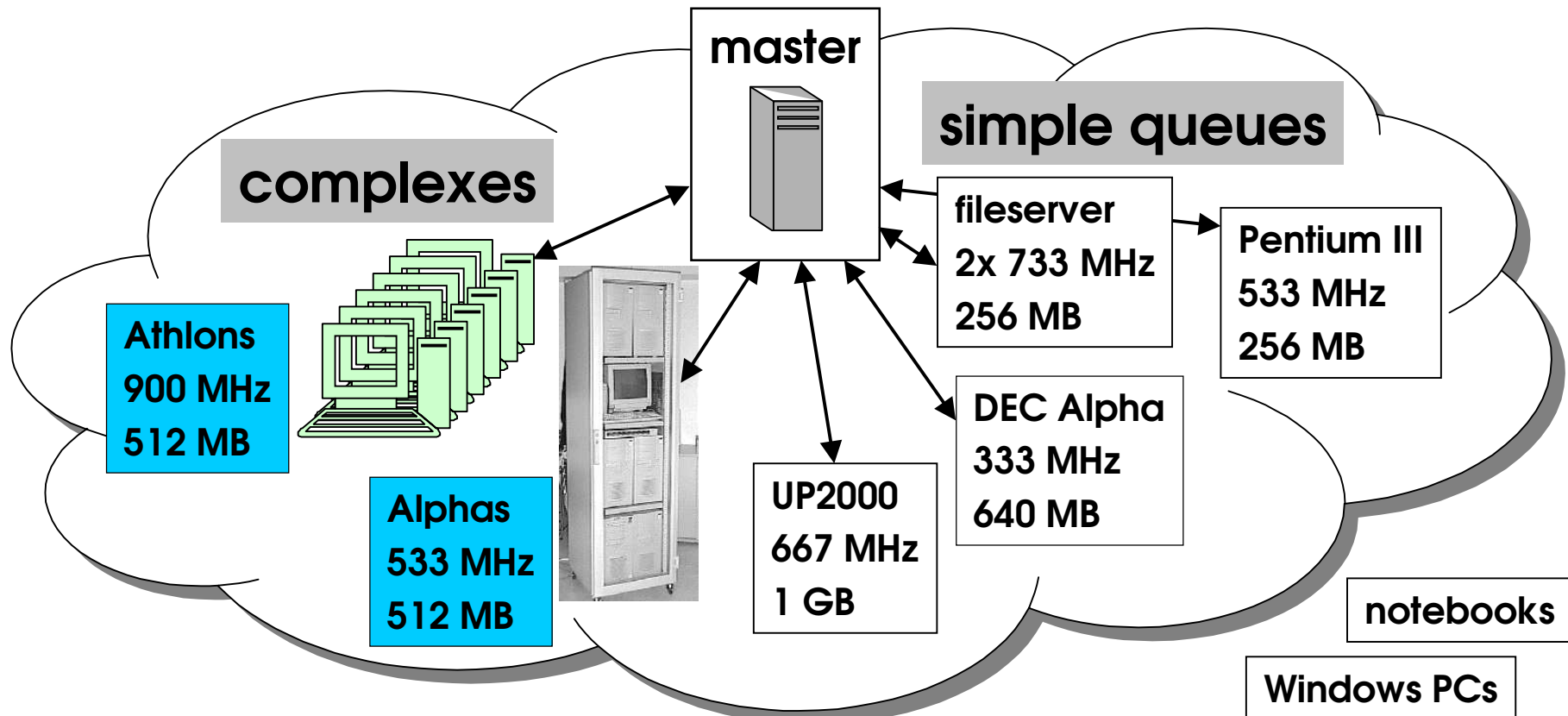
- **Centralized data storage on the fileserver**
 - /home: 80 GB on SCSI hardware RAID 5
 - /scr: 110 GB on IDE software RAID 0 (2 disks)
 - /scrib: 110 GB on IDE software RAID 0 (2 disks)
- **Daily backups on DAT tapes (autoloader)**
- **Weekly virus scans (reports by email)**
- **Access**
 - NFS mounting from all clients
 - Samba for access from Windows PCs
 - no data transfer (e.g. FTP) required



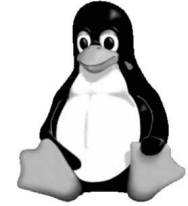
Queueing system



- all Linux/Unix machines use NIS/yellow pages for authentication
- all machines have NFS access to the fileserver
- all Linux/Unix machines form a DQS cluster



User interface



```
scholz@a01:~$ qstat -f
Queue Name
-----
a00
  forster  010_003      batch  4698  0:6  r  1.09  er  UP
a01
  scholz   d1adx1      batch  4704  0:3  r  1.22  er  UP
a02
  scholz   d02ad13    batch  4688  0:1  r  1.15  er  UP
a03
  forster  005_003    batch  4699  0:7  r  1.00  er  UP
a04
  scholz   d02adx1    batch  4695  0:2  r  1.04  er  UP
a05
a10
  scholz   d0505x1    batch  4705  0:4  r  1.00  er  UP
atp330
fs
fx01
  forster  100_002    batch  4682  0:2  r  1.01  er  UP
fx02
  forster  001_001    batch  4700  0:8  r  1.00  er  UP
fx03
  forster  100_001    batch  4683  0:3  r  1.10  er  UP
fx04
  forster  001_003    batch  4701  0:9  r  1.00  er  UP
fx05
  forster  001_003    batch  4701  0:9  r  1.00  er  UP
fx06
  forster  001_003    batch  4701  0:9  r  1.00  er  UP
up01
  forster  010_001    batch  4696  0:4  r  2.02  er  UP
  forster  005_001    batch  4697  0:5  r  2.02  er  UP
scholz@a01:~$
```

Athlons
900 MHz
512 MB

Pentium III
533 MHz
256 MB

DEC Alpha
333 MHz
640 MB

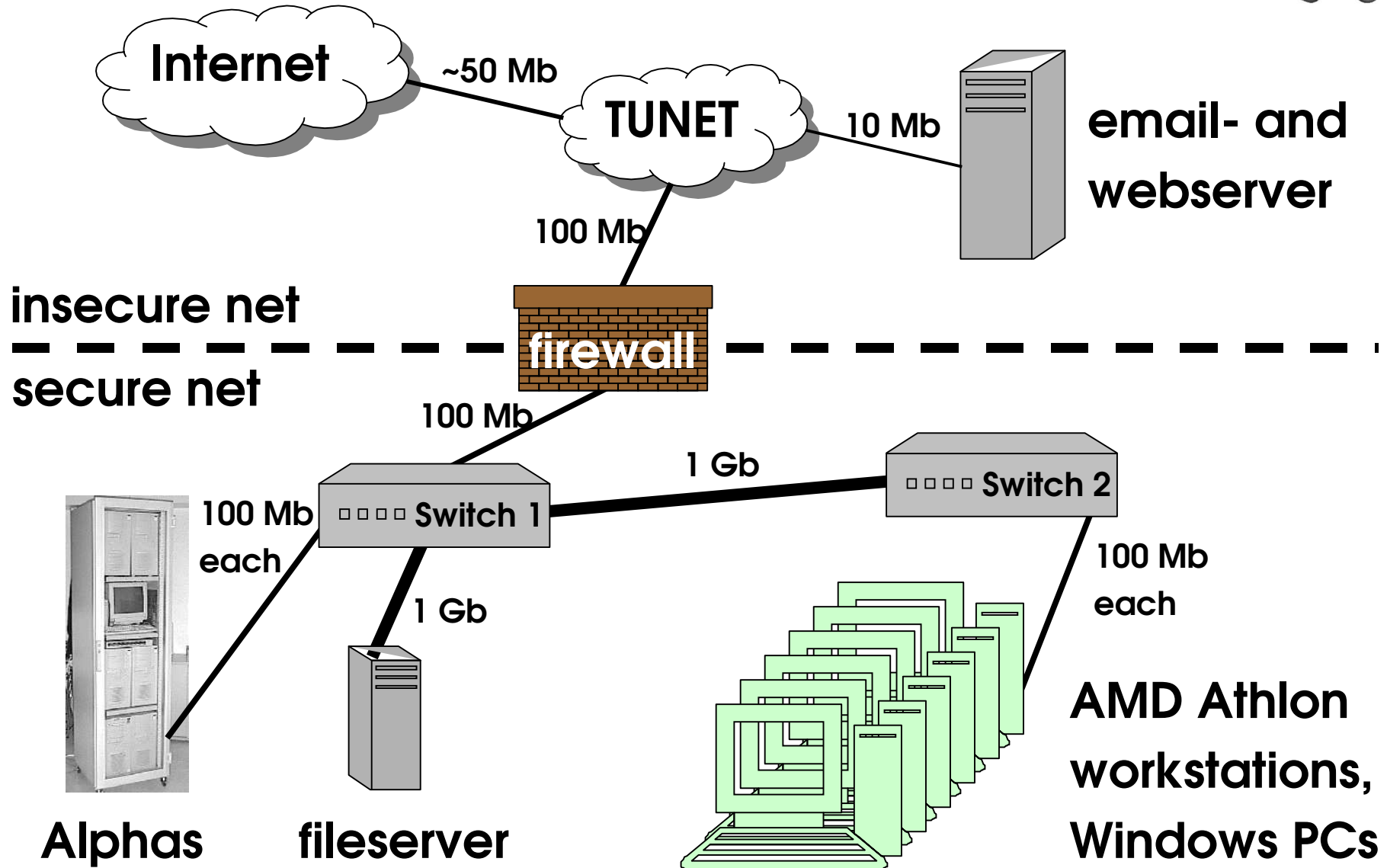
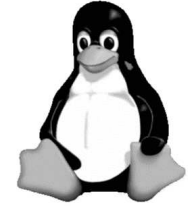
fileserver
2x 733 MHz
256 MB

Alphas
533 MHz
512 MB

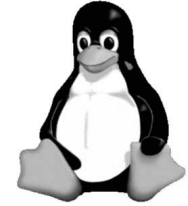
UP2000
667 MHz
1 GB

- Required commands for users:
qsub, qstat, qdel

Network layout

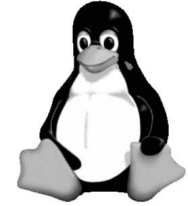


Remote access

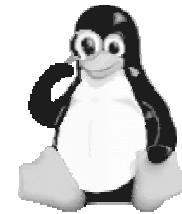


- **Threats:**
 - password sniffing (unauthorized access)
vulnerable: http, telnet, ftp, pop, imap
 - listening in on non-encrypted communication
- **Defense:**
 - one time passwords (s/key)
 - encrypted authentication
 - encrypted communication channels
 - detailed logging
- **Tools:**
 - SSH (replaces telnet, ftp; tunneling)
Windows clients available
 - SSL (https, pops, imaps)
built into Outlook, Netscape, etc.

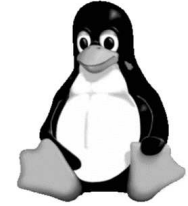
Security



- Possible security holes:
 - NIS/yellow pages for authentication
 - NFS
 - FTP
- Firewall is an absolute requirement
- Secure encrypted communication with SSH
- Continuous security auditing
- Installation of patches, updates
- Security alerts from
 - CERT: Carnegie Mellon Software Engineering Inst.
 - SANS, Incidents.org, dshield.org, etc.

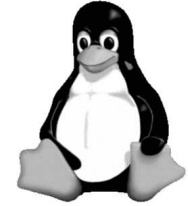


Management



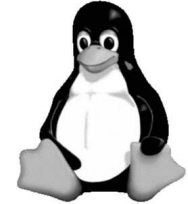
- **Client setup with RedHat kickstart installation or harddisk cloning (even full VMware installation)**
- **Beowulf tools**
- **Automation with shell scripts and cron jobs**
- **Cluster management**
 - check for free hard disk space
 - weekly report of hard disk usage
 - check for unreachable machines
- **Security related**
 - backups (data, configuration, emails)
 - virus scanning (data and emails)
 - update of antivirus databases
 - check for unauthorized logins
- **notification with SMS to GSM phones or to pagers**

Summary



- **Linux is a proven server OS**
- **Linux is ready for the desktop**
- **integrates easily in a heterogeneous network**
- **simple clustering**
- **lean installation**
- **best use of available resources**
- **efficient maintenance**

References



1. samba - opening windows to a wider world
<http://www.samba.org/>
2. SuSE Linux AG
<http://www.suse.de/>
3. Red Hat, Inc.
<http://www.redhat.com/>
4. API NetWorks Inc.
<http://www.alpha-processor.com/products/up2000-board.shtml>
5. DQS - Distributed Queueing System
<http://www.scri.fsu.edu/~pasko/dqs.html>
6. The Beowulf Project
<http://www.beowulf.org>
7. The Beowulf Underground
<http://buweb.parl.clemson.edu/>
8. SSH Communications Security
<http://www.ssh.fi/>
9. MindTerm - pure java ssh Client
<http://www.mindbright.se/>
10. OpenSSH
<http://www.openssh.org/>
11. PuTTY: A Free Win32 Telnet/SSH Client
<http://www.chiark.greenend.org.uk/~sgtatham/putty.html>
12. WinSCP - secure data transmission
<http://winscp.vse.cz/>
13. Secure iXplorer - Windows Front End for PSCP
<http://www.i-tree.org/ixplorer.htm>
14. SSH Clients for other operating systems
<http://www.openssh.org/windows.html>
<http://www.openssh.org/unix.html>
<http://opensores.thebunker.net/pub/mirrors/ssh-faq/ssh-faq-2.html#ss2.1>
15. sendmail.org - Masquerading and Relaying
<http://www.sendmail.org/m4/masquerading.html>
16. CERT Coordination Center
<http://www.cert.org/>
17. SANS Global Incident Analysis Center
<http://www.sans.org/>