Vollautomatische Installationen mit FAI

Nürnberg, April 2011

Thomas Lange, Universtät zu Köln

lange@informatik.uni-koeln.de

Agenda

- Warum automatisch installieren?
- Crashtest
- Wie funktioniert FAI?
- Leistungsdaten
- Nachteile der Automatisierung
- Erfahrungen mit FAI

- whoami
  - Diplominformatiker, Uni Bonn
  - Systemadministrator seit über 19 Jahren
  - SunOS 4.1.1 auf SPARC
  - Solaris Jumpstart
  - 1999 erstes 16 Knoten Cluster (Dual PII 400MHz)
  - FAI seit 10+ Jahren
  - Debian Entwickler seit 2000
  - Vorträge und Tutorials auf zahlreichen Konferenzen:

    Linux Kongress, Linuxtag, DebConf, SANE, LCA, FOSDEM, SUCON, CeBit, OSDC, UKUUG

Manuelle Installation?

Wer möchte diese Rechner per Hand installieren?

168 IBM HS20 Blades, 2x2.8 GHz

90 dual Itanium 2, 900Mhz

www.centibots.org

Was ist ein Linux Rollout?

- Geplante Installation
- Aufsetzen von Betriebssystem und Anwendungen
- Installation und Konfiguration
- Zentrale Verwaltung und Steuerung

Weitere Ziele eines Linux Rollouts

- Heterogene Konfigurationen unterstützen
- Automatische Dokumentation
- Inventarisierung
- Disaster recovery
- Computer Infrastruktur bauen

Manuelle Installation?

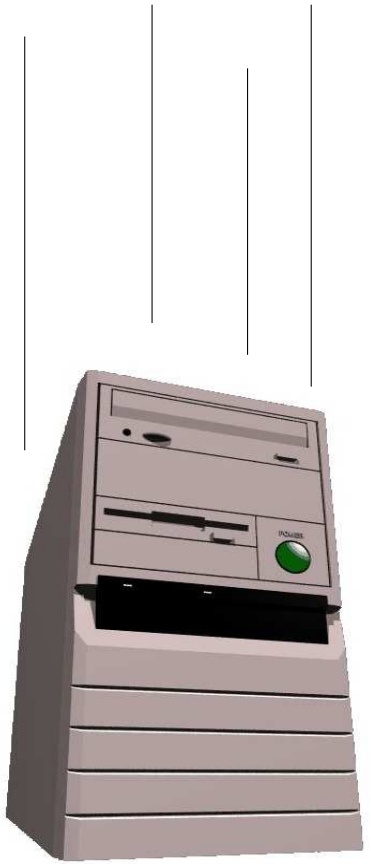
180 dual AMD MP2200, Max Planck Institute for Gravitational Physics

Wert eines Computers

- Was beinhalten ihre Rechner?
  - Kundendaten
  - Dienste
  - Applikationen
  - Eigenes Know-How
- Was passiert, wenn ihre Rechner einen Tag lang nicht laufen?
- Eine gute Computerinfrastruktur ist so wichtig wie ...
- Wie sichern Sie diese Werte?
- Ist damit wirklich alles gesichert?

Der Crashtest

- Wählen Sie zufällig einen Rechner (ohne Backup vorher)
- Werfen sie den Rechner aus dem 10.Stock (oder `dd if=/dev/zero of=/dev/hda`)



- Stellen Sie alle Arbeit des Sysadmin innerhalb von 10 Minuten wieder her
- Schaffen Sie das?

Fakten, die oft übersehen werden

- Gut laufende Rechner sind ihr Kapital
- Backup der Daten ist nur ein Teil
- Haben Sie eine Kopie ihres Sysadmins?

- **Alles automatisieren!**

Manuelle Installation

- Dauert viele Stunden
- Dokumentation fehlt, Reproduzierbarkeit?
- Viele Fragen
- Wiederholende Arbeit ist stupide => Fehler
- Jede Installation ist ungewollt einzigartig
- ”No simple sysadmin task is fun more than twice”
- **Eine Installation per Hand skaliert nicht !**

Warum voll automatisch?

- Garantiert identische Installationen
- Automatische Dokumentation
- Heterogene Hardware und unterschiedliche Konfigurationen
- Hilft schnell nach Hardwaredefekt
- Parallele Installationen
- Spart sehr viel manuelle Arbeit (= Zeit = Geld)
- Macht mehr Spaß

Was ist FAI ?

- FAI macht alles, was ihr Systemadministrator zu tun hat, bevor der Benutzer das erste Mal auf einem neuen Rechner arbeiten kann
- Serverbasiertes Tool
- Skriptgesteuert
- Installiert und konfiguriert das OS und Anwendungsprogramme
- Kein Master Image
- Modular durch Klassensystem
- Erweiterbar und flexibel durch hooks
- Es kann die Installation nicht planen :-(, aber
- **Plane deine Installation und FAI installiert deinen Plan! :-)**

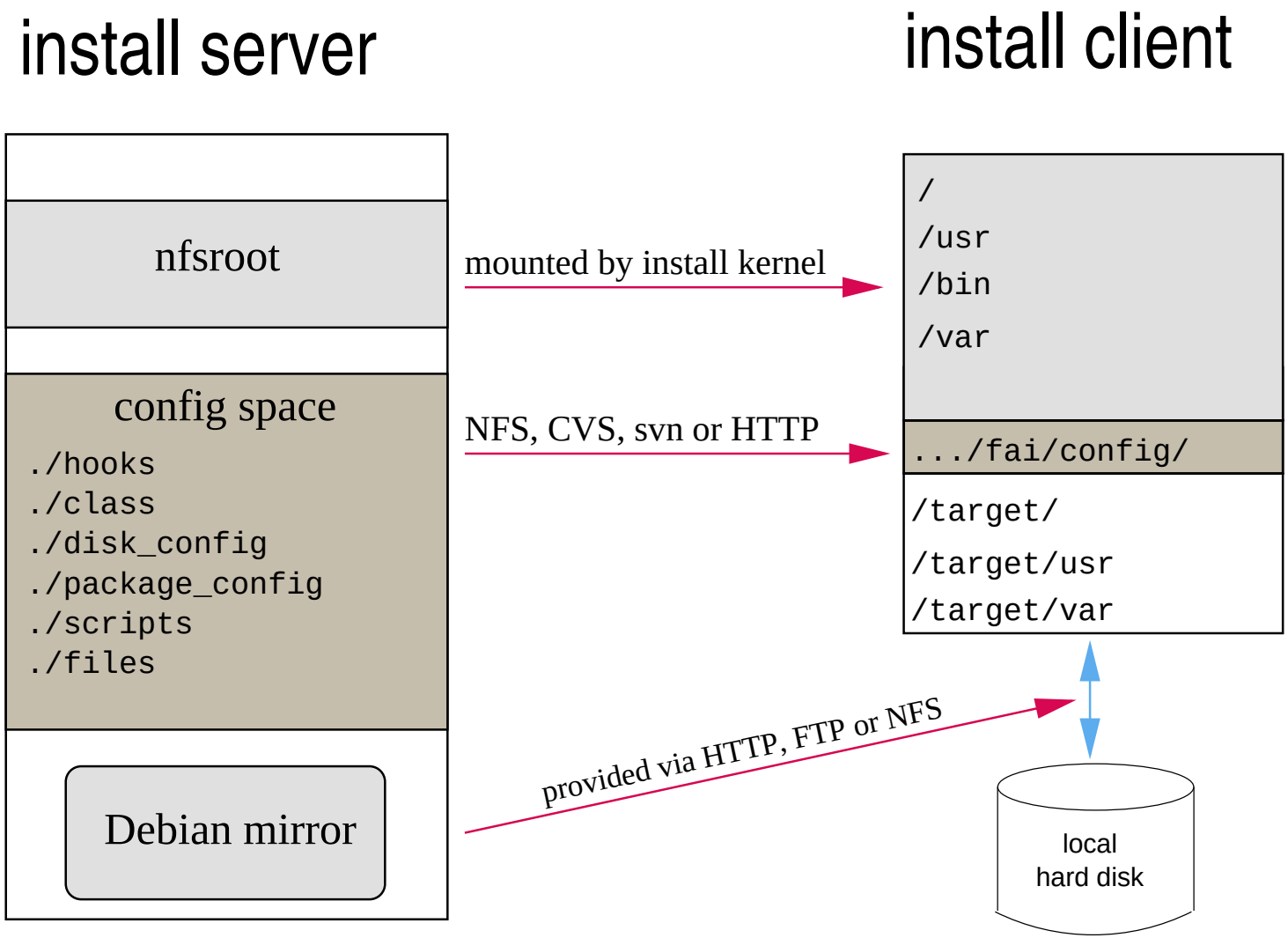
Was braucht FAI?

- Installserver mit DHCP, NFS und TFTP
- Client bootet via PXE, CD-ROM, USB Stick
- Lokaler Spiegel von Debian (NFS, FTP oder HTTP)
- Plattenplatz auf dem Server:

FAI Pakete	1 MB	Skripte, Konfigurationdateien
nfsroot	450 MB	erzeugt mit `make-fai-nfsroot`
Debian Spiegel	34 GB	Debian 6.0 (squeeze, nur i386)

- Alle Install Clients nutzen die gleichen Verzeichnisse
- **Konstanter Plattenplatz**

Wie funktioniert FAI ?



- Die Konfiguration liegt auf dem Install server
- Die Installation läuft auf dem Klienten

Ablauf einer Installation I

- Plane deine Installation!
- Booten via PXE und Kernel mit initrd via TFTP holen
- Rechner startet als Diskless Client
- Hardwareerkennung und Kernel Module laden

Ablauf einer Installation II

- Klassen und Variablen definieren
- Festplatten partitionieren
- Dateisysteme erzeugen und mounten
- Software Pakete installieren
- Betriebssystem und Anwendungen konfigurieren
- Protokolldateien lokal und auf Install Server speichern
- Neu installiertes System booten

Das Klassenkonzept

- Ein Rechner gehört zu mehreren Klassen
- Priorität von niedrig nach hoch
- Beispiel: `DEFAULT FAIBASE GRUB GNOME demohost LAST`
- Alle Teile der Installation nutzen das Klassenkonzept
- Konfiguratitonsdateien werden anhand der Klassennamen ausgewählt
- `fcopy` zum Kopieren von Templates

- Erfahrener Admin kreiert die Klassen
- Junior Admin ordnet die Klassen den Rechnern zu
- PC installiert sich selber

Verzeichnisse im Config Space

```
|-- class
|   |-- 10-base-classes
|   |-- 20-hwdetect.source
|   |-- 50-host-classes
|   |-- FAIBASE.var
|   `-- GERMAN.var
|-- disk_config/
|   |-- FAIBASE
|   |-- SMALL_IDE
|   `-- foobar04
|-- debconf
|   `-- FAIBASE
|-- package_config/
|   |-- FAIBASE
|   |-- DEBIAN_DEVEL
|   |-- GERMAN
|   |-- GNOME
|   `-- server07
```

Beispiel: `.../class/10-base-classes`:

```
#! /bin/sh

dpkg --print-architecture | tr a-z A-Z       # I386

case $HOSTNAME in
    demohost)
        echo "FAIBASE DHCPC DEMO" ;;
    gnomehost)
        echo "FAIBASE DHCPC DEMO XFREE GNOME";;
esac

case $IPADDR in
    134.95.9.*) echo "CS_KOELN NET_9" ;;
esac

ifclass I386 && echo "GRUB"

lspci | grep -q MATROX || echo "MATROX"
```

Beispiel: `.../class/FAIBASE.var`:

```
FAI_ALLOW_UNSIGNED=1

KEYMAP=de-latin1-nodeadkeys
UTC=yes
TIMEZONE=Europe/Berlin

ROOTPW='$1$kBnWcO.E$djxB128U7dMkrltJHPf6d1'
LOGUSER=fai

flag_initial=1   # used by setup-storage
```

- Eigene Variablen möglich
- Die Konfigurationsskripte in `.../scripts/*` nutzten diese Variabeln

Neue Plattenpartitionierung

Beispiel: `.../disk_config/FAIBASE`:

```
disk_config disk1     preserve:9 fstabkey:uuid

primary  /      300-900    ext4  rw,errors=remount-ro
logical  swap   1G         swap  rw
logical  /usr   2G-4G      ext4  noatime,rw
logical  /var   1G-2G      ext4  rw   createopts="-L var -m 5"
logical  /tmp   50-1000    ext4  rw     tuneopts="-c 0 -i 0"
logical  /home  5G-        ext4  defaults
```

- Filesysteme: ext2/3/4, vfat, xfs, ReiserFS, NTFS

```
disk_config disk1

primary /boot 20-100      ext4       rw
primary  swap 1024        swap       sw
primary /     2000-4000   ext4       rw,acl,user_xattr
logical  -    0-          - -
logical  -    0-          - -
logical  -    0-          - -
logical  -    0-          - -

disk_config raid
raid1    -    disk1.5,disk1.7    - -
raid1    -    disk1.6,disk1.8    - -

disk_config lvm
vg  volg1  md0,md1
volg1-usr  /usr          2048 ext4  rw createopts="-O dir_index,resize_inode"
volg1-var  /var          600  ext4  rw createopts="-O dir_index,resize_inode"
volg1-hl   /home/local   4096 ext4  rw,acl,user_xattr,noexec,nosuid,nodev
volg1-es   /export/sites 2048 ext4  rw createopts="-O none"
volg1-v    /vservers     2048 ext4  rw createopts="-O ^dir_index,^resize_inode"
```

Beispiel: `.../package_config/BEOWULF`:

```
# packages for Beowulf clients

PACKAGES install BEOWULF_MASTER
gmetad apache

PACKAGES aptitude
fping jmon ganglia-monitor
rsh-client rsh-server rstat-client rstatd rusers rusersd

dsh update-cluster-hosts update-cluster etherwake

lam-runtime lam4 lam4-dev libpvm3 pvm-dev mpich
scalapack-mpich-dev
```

- Aktionen `aptitude, apt-get, smart, rpm, urpmi, y2pmsh, yast, yum, zypper`
- Abhängigkeiten innerhalb der Pakete werden aufgelöst

Verschnaufpause

Top500: 58th in 6/2008, 1340 nodes, 5376 cores, Xeon 2.4 GHz

Max Planck Institute for Gravitational Physics

Verzeichnisse im Config Space

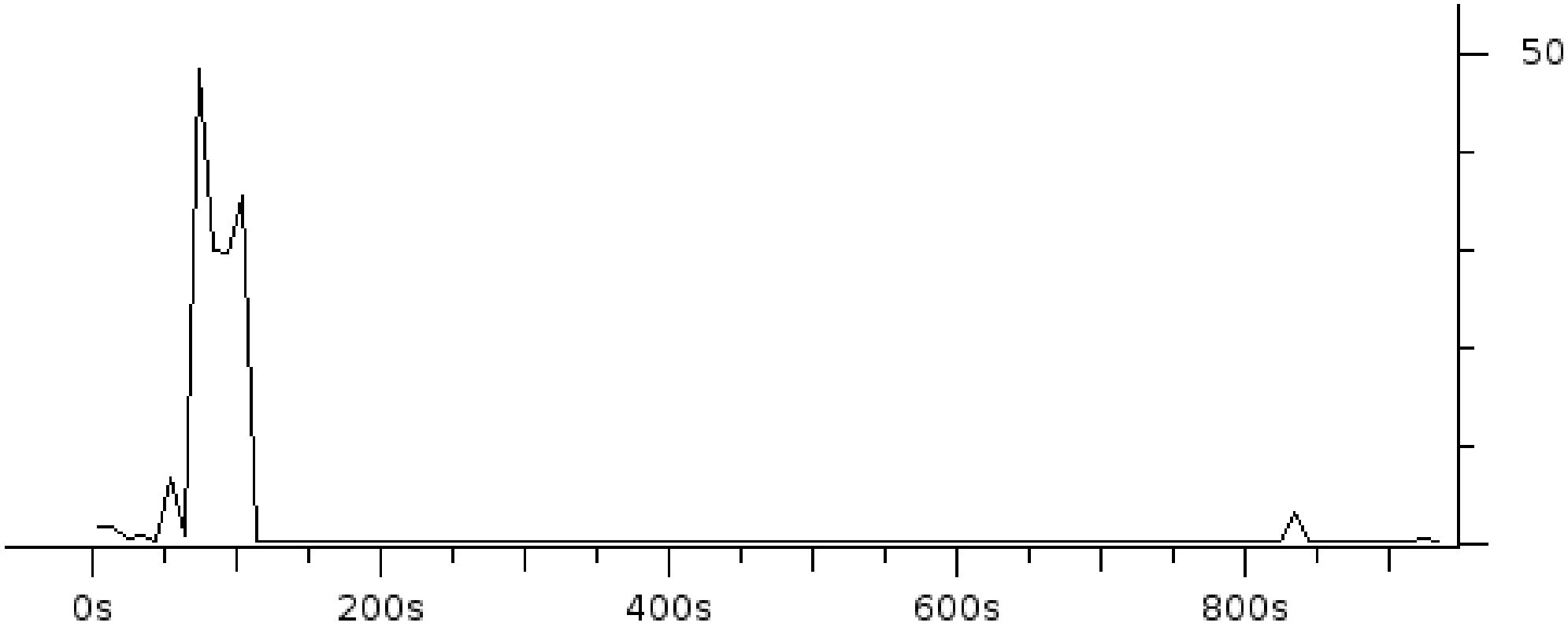
```
|-- scripts/
|   |-- BOOT
|   |-- FAIBASE/
|   |   |-- 10-misc                     Bourne shell script
|   |   |-- 30-interface                Bourne shell script
|   |   `-- 40-misc                     /usr/bin/cfengine script
|   |-- DEMO/
|   |   |-- 10-misc                     Bourne shell script
|   |   `-- 30-demo                     /usr/bin/cfengine script
|   `-- demohost
`- files/
    `-- etc/
        `-- X11/
            `-- xorg.xonf/          fcopy /etc/X11/xorg.conf
                |-- FAIBASE
                |-- MATROX
                `-- demohost
```

Konfigurationsskripte

```
# create NIS/NONIS config
fcopy -M /etc/nsswitch.conf /etc/host.conf
fcopy -i /etc/ypserv.securenets  # only for yp server
ifclass NONIS && rm -f $target/etc/defaultdomain
if ifclass NIS; then
    echo $YPDOMAIN > $target/etc/defaultdomain
    rm -f $target/etc/yp.conf
    for s in $YPSRVR; do
        ainsl -av $target/etc/yp.conf "ypserver $s"
        # don’t do this! # echo "ypserver $s" >> $target/etc/yp.conf
    done
fi

ainsl -v $target/etc/fstab "$bserver:/usr/local /usr/local nfs ro 0 0"

fcopy -M /etc/X11/xorg.conf
```

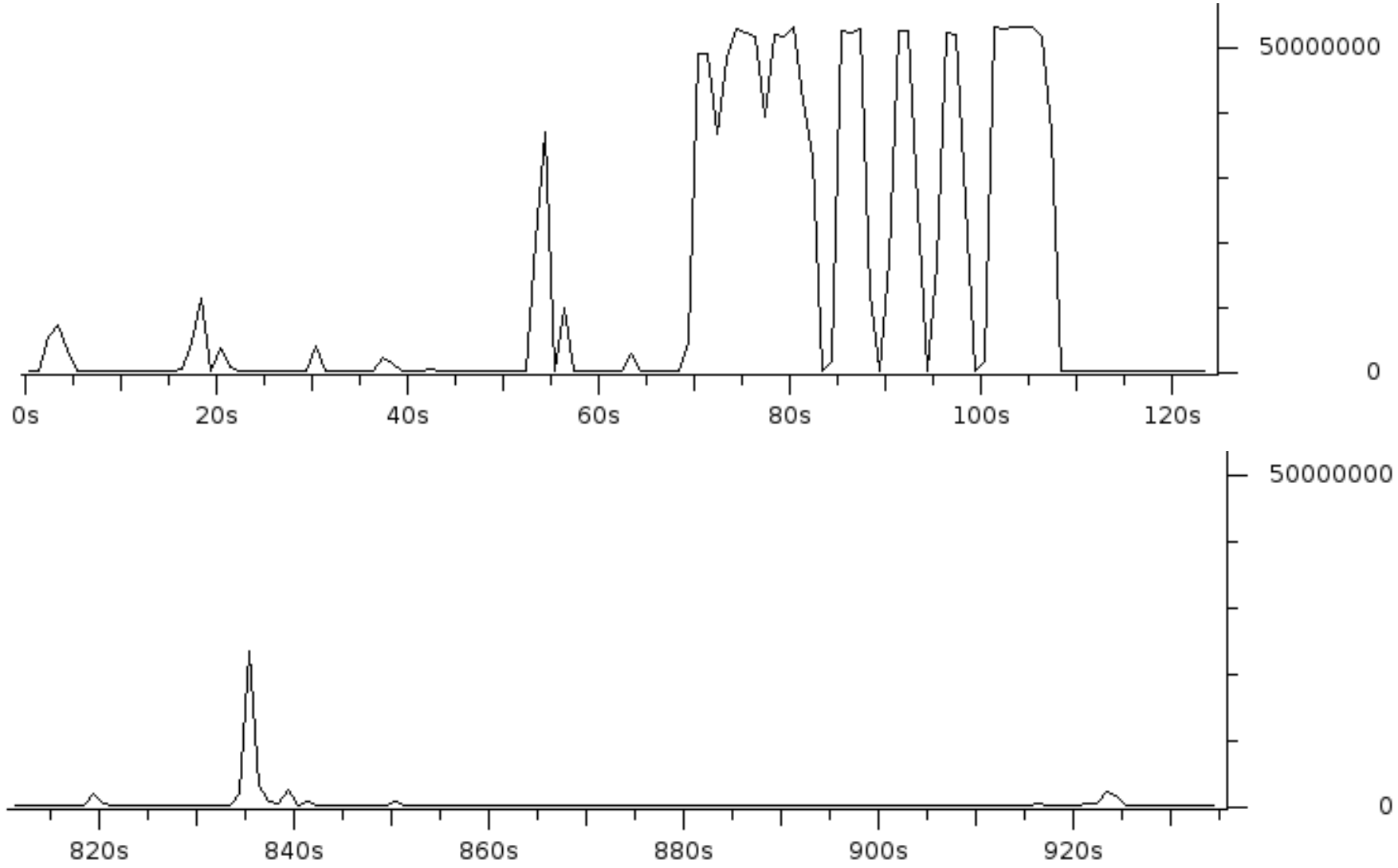
```
files:
   any::
      ${target}/dev include=fd* mode=666   action=fixall r=1

editfiles:
   any::
   { ${target}/etc/fstab
         AppendIfNoSuchLine "none  /proc/bus/usb  usbdevfs  defaults"
         AppendIfNoSuchLine "/dev/fd0  /floppy  auto  users,noauto 0 0"
   }
   { ${target}/etc/inittab
      ReplaceAll "/sbin/getty" With "/sbin/getty -f /etc/issue.linuxlogo"
   }
 TERMINAL_CLIENT::
   { ${target}/etc/inetd.conf
     HashCommentLinesContaining "in.rlogin"
   }
```

- Rechner Core2duo, GBit LAN, Debian squeeze, amd64
- Installation dauert 15 min



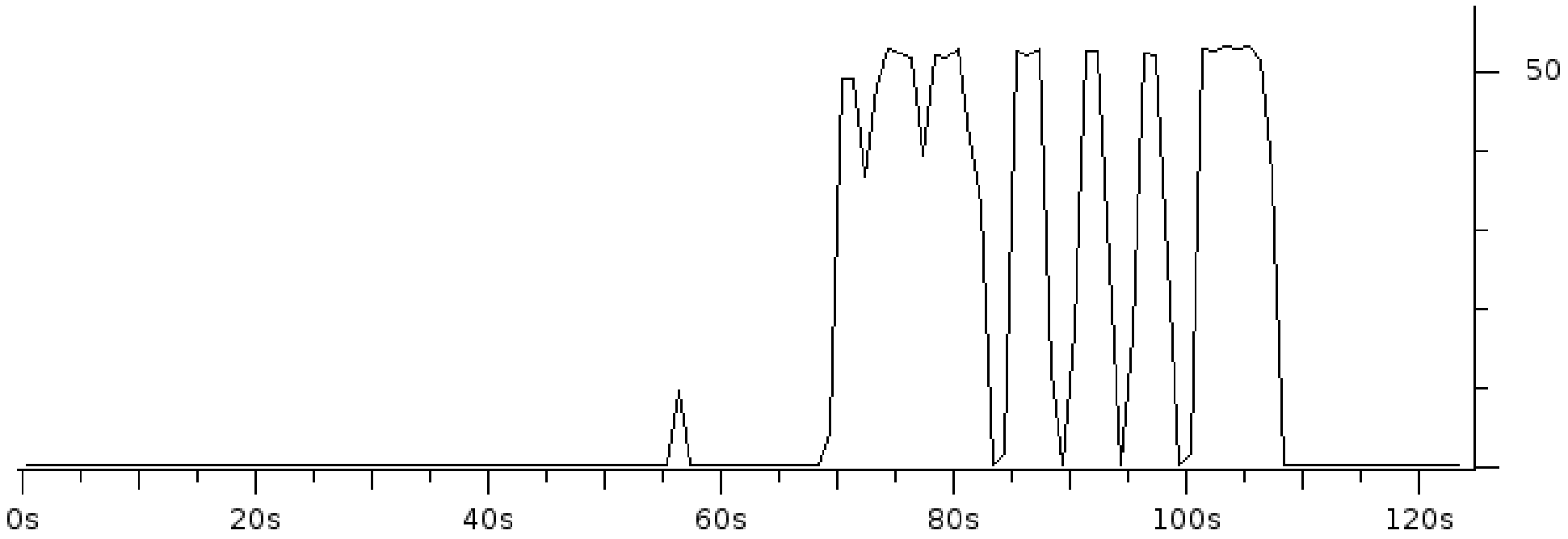
Netzwerkverkehr während einer Installation

Netzwerk II



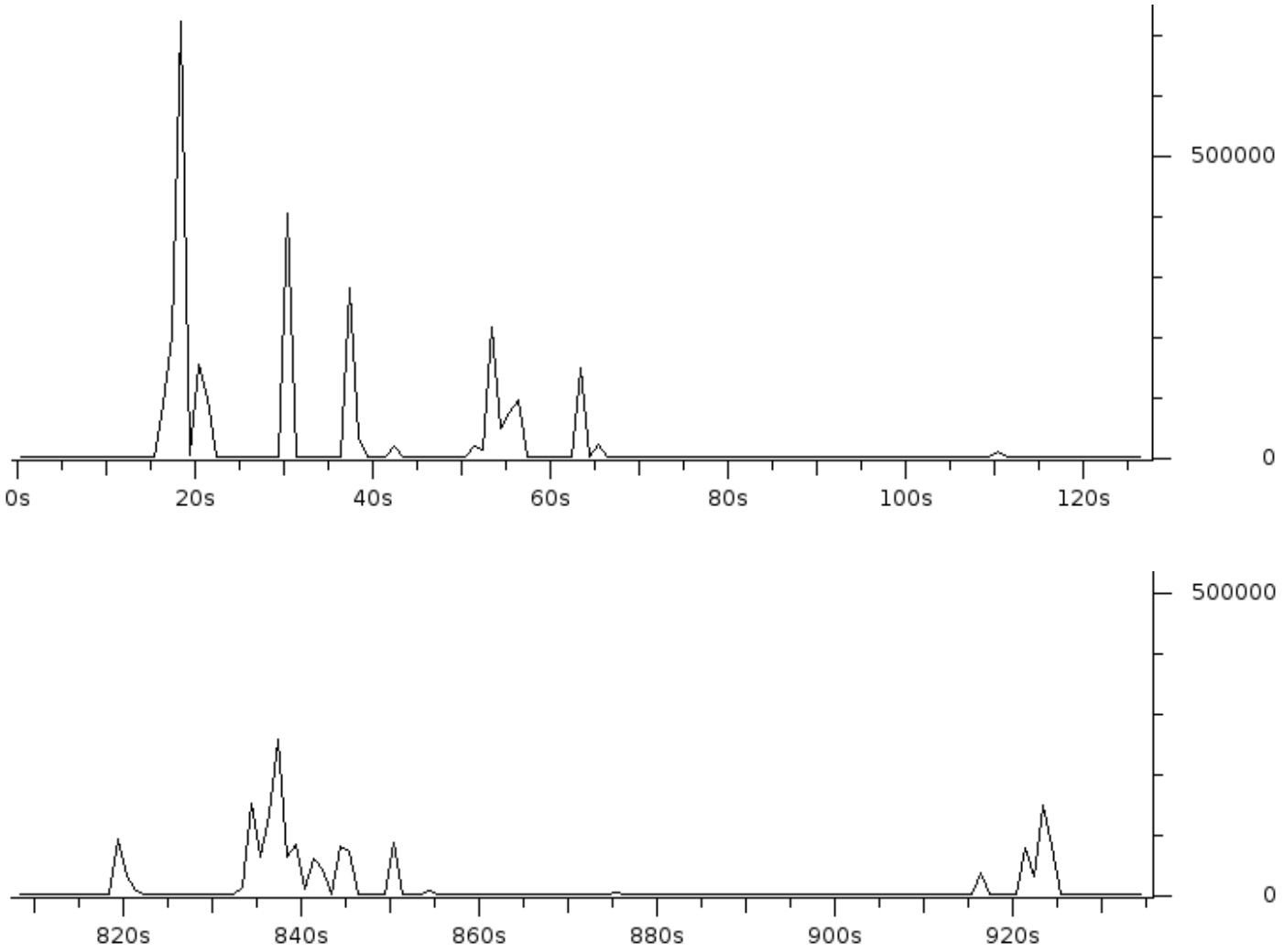
Gesamter Datenverkehr in der Anfangs- und Endphase

HTTP Pakete: zw. 70s-110s werden ca. 1.3GB Daten übertragen mit 36MB/s



HTTP Pakete

NFS Pakete in der Anfangs- und Endphase



Network traffic I

- squeeze 64bit, 2.6.32
- 4.2GB software installed, package download 1361 MB
- kernel and initrd = 14 MB data (TFTP)
- base.tgz 50 MB data (extracted 134 MB) (NFS)
- network traffic: RX: 1.45 GB TX: 7.96 MB (without kernel,initrd)

```
IP:        1.46 GB          100.00%
TCP:       1.44 GB           99.01%

HTTP:      1.34 GB           92.18%
NFS:         100   MB         6.76%
TFTP:         14.6 MB         0.98%
SSH:           1   MB         0.07%

DOMAIN: 77kB       NTP: 10kB
```

Network traffic II

- Same as above
- 467 MB software installed, package download 61.6 MB
- Network traffic: RX: 158.6 MB TX: 3.2 MB (without kernel,initrd)

```
IP:         174 MB            100.00%
TCP:        159 MB             91.57%

HTTP:        70 MB             40.29%
NFS:         89 MB             51.18%
TFTP:        14.6   MB          8.42%
SSH:          0.158 MB          0.09%

NTP: 10kB      DOMAIN: 4kB
```

- NFS can be reduced to 39 MB
- Use base.tar.xz (31 MB) and receive via HTTP (instead of NFS)

- 64 bit squeeze, 2.6.32 kernel, dpkg 1.15.8.7, noatime
- ext4 uses barrier=1 as default
- 1358 MB package download, 4.2 GB software installed
- Ramdisk means `/var/lib/dpkg` in RAM

Machine A:

- Core2duo 2GHz, 2GB RAM, 2,5 zoll harddisk, 120GB 5400rpm
- 1358 MB package download (needs 40-50s), 28-34 MB/s
- misc time for installation 80-140s, 10GB mkfs

Zeiten Rechner A

Installationszeiten (task instsoft) ohne Download der Pakete

FS	RAM disk	unsafe-io	misc	Time	Scale
ext4	N	N	nodelalloc	1865	
ext4	N	N		1821	2.62
ext4	N	Y		1736	
ext4	Y	N		887	1.28
ext4	N	N	barrier=0	802	
ext4	Y	Y		773	1.12
ext4	Y	Y	barrier=0	693	1
ext3	N	N	barrier=0	926	
ext3	Y	Y	barrier=0	783	

Installationszeiten

Host, RAM	Software	Zeit
Core i7, 3.2 GHz, 6GB	4.3 GB	7 min
Core i7, 3.2 GHz, 6GB	471 MB	77 s
Core2duo, 2 GHz, 2GB	4.3 GB	17 min
Core2duo, 2 GHz, 2GB	471 MB	165 s
Pentium 4, 3 GHz, 1GB	2200 MB	10 min
Pentium 4, 3 GHz, 1GB	1100 MB	6 min
Pentium 4, 3 GHz, 1GB	300 MB	105 s

Knoten	Sekunden
1	337
5	340
10	345
20	379

12% mehr Zeit bei 20 Rechnern.

Noch ein Beispiel

356 opterons, 80 xeons, Top500 in 11/2005,
Trinity Centre for High Per formance Computing, Dublin

FAI Benutzer

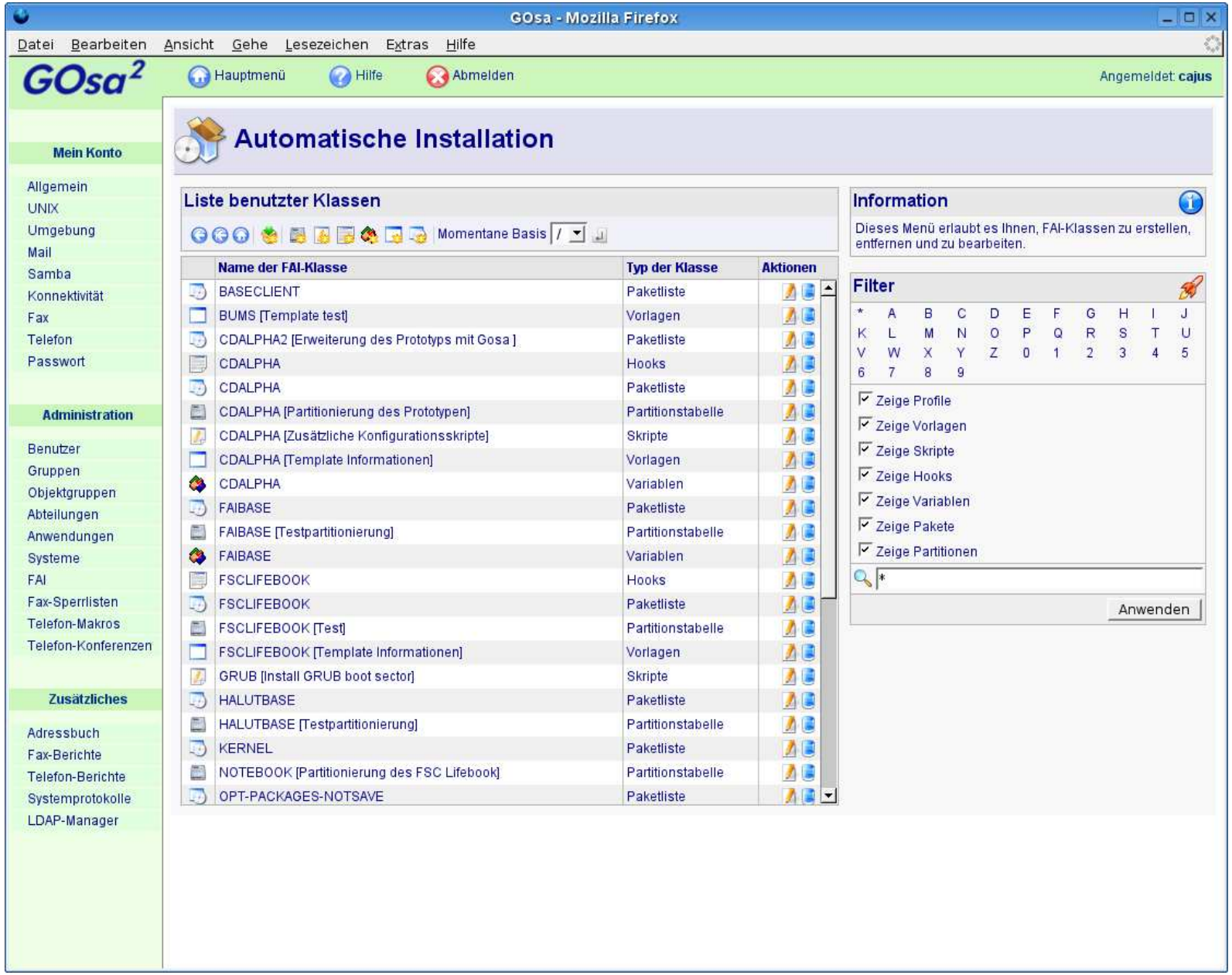
- City of Munich, ∼5500, 14.000 hosts planned
- Zivit, 260 hosts on two IBM z10 EC mainframes
- Archive.org, 200+ hosts
- XING AG, 300-400 hosts
- Opera Software, ∼300 hosts
- Stanford University, 450 hosts
- MIT Computer science research lab, 200 hosts
- The Welcome Trust Sanger Institute, 540 hosts
- Deutsches Elektronen-Synchrotron, 273 hosts
- Mobile.de, ∼600 hosts
- Electricité de France (EDF), 1500 hosts
- BUF, digital visual effects company, 1000 hosts
- ETH Zurich, systems group, ∼300 hosts
- High Performance Computing Center North, HPC2N, two clusters with a total of 310 hosts
- NETWAYS, Netcologne, MPI Meteorologie, DESY, Genua, taz, thomas-krenn.com

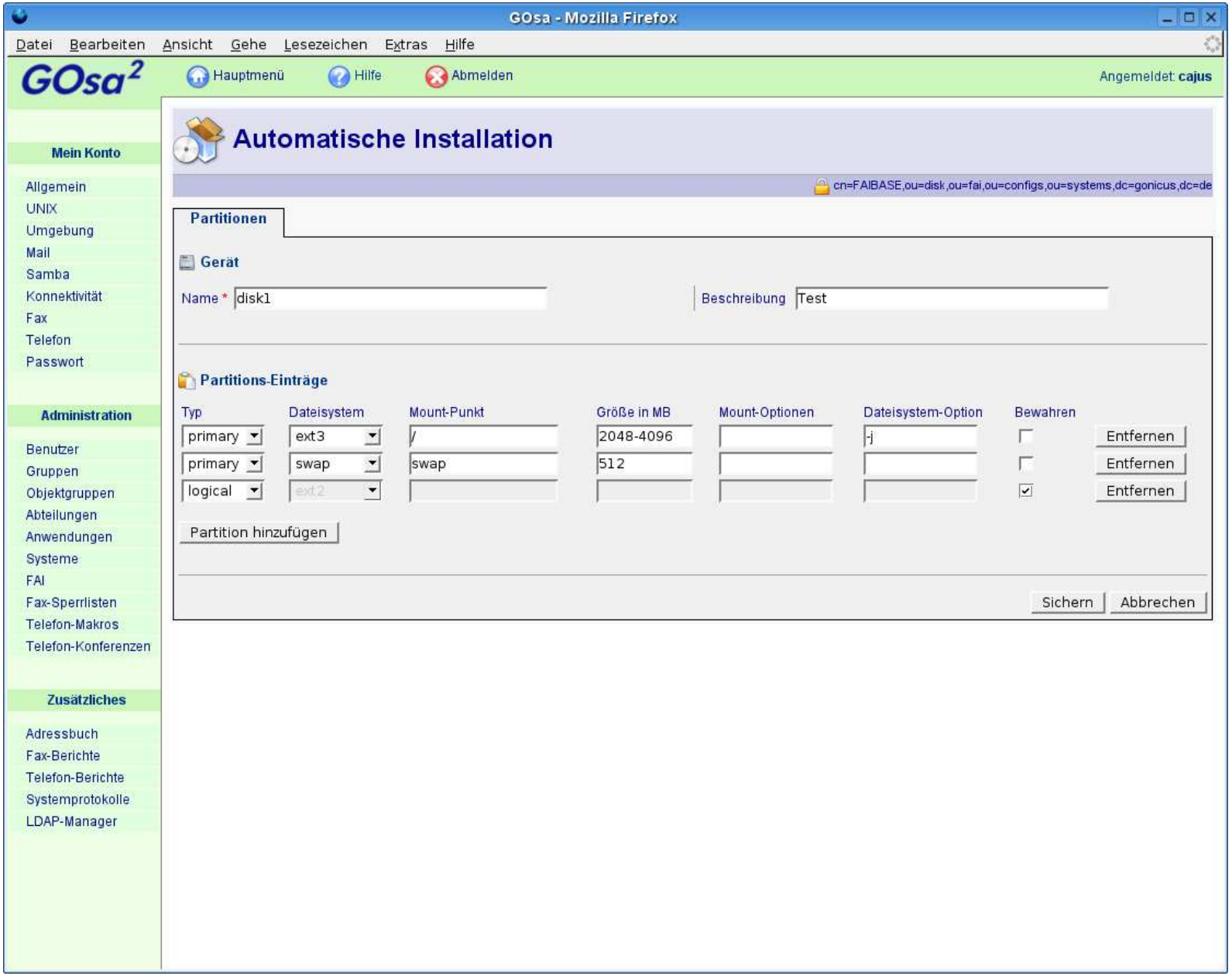
Nachteile der Automatisierung

- Kann oder soll alles automatisiert werden?
- Fehler werden auch verteilt
- Weniger aber höher qualifiziertes Personal notwendig
- Man muss erstmal Zeit und Arbeit investieren
- Bereitschaft für Veränderungen?
- Sysadmin wird zum sauberem Arbeiten gezwungen
- Manuelle Änderungen an einzelnen Rechner sind verboten!

Weiteres in FAI

- Andere Distributionen, auch RPM
- Virtuelle Rechner installieren
- Aufsetzen von chroot (z.B für Live CD’s, grml)
- Multiarchitekur Server für i386 und amd64
- Softupdates
- GOsa


GOsa - Mozilla Firefox
Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe
GOsa²
Hauptmenü
Hilfe
Abmelden
Angemeldet: cajus
Mein Konto
Allgemein
UNIX
Umgebung
Mail
Samba
Konnektivität
Fax
Telefon
Passwort
Administration
Benutzer
Gruppen
Objektgruppen
Abteilungen
Anwendungen
Systeme
FAI
Fax-Sperrlisten
Telefon-Makros
Telefon-Konferenzen
Zusätzliches
Adressbuch
Fax-Berichte
Telefon-Berichte
Systemprotokolle
LDAP-Manager
Automatische Installation
Liste benutzter Klassen
Momentane Basis /
Name der FAI-Klasse
Typ der Klasse
Aktionen
BASECLIENT Paketliste
BUMS [Template test] Vorlagen
CDALPHA2 [Erweiterung des Prototyps mit Gosa] Paketliste
CDALPHA Hooks
CDALPHA Paketliste
CDALPHA [Partitionierung des Prototypen] Partitionstabelle
CDALPHA [Zusätzliche Konfigurationsskripte] Skripte
CDALPHA [Template Informationen] Vorlagen
CDALPHA Variablen
FAIBASE Paketliste
FAIBASE [Testpartitionierung] Partitionstabelle
FAIBASE Variablen
FSCLIFEBOOK Hooks
FSCLIFEBOOK Paketliste
FSCLIFEBOOK [Test] Partitionstabelle
FSCLIFEBOOK [Template Informationen] Vorlagen
GRUB [Install GRUB boot sector] Skripte
HALUTBASE Paketliste
HALUTBASE [Testpartitionierung] Partitionstabelle
KERNEL Paketliste
NOTEBOOK [Partitionierung des FSC Lifebook] Partitionstabelle
OPT-PACKAGES-NOTSAVE Paketliste
Information
Dieses Menü erlaubt es Ihnen, FAI-Klassen zu erstellen, entfernen und zu bearbeiten.
Filter
* A B C D E F G H I J
K L M N O P Q R S T U
V W X Y Z 0 1 2 3 4 5
6 7 8 9
Zeige Profile
Zeige Vorlagen
Zeige Skripte
Zeige Hooks
Zeige Variablen
Zeige Pakete
Zeige Partitionen
*
Anwenden


GOsa - Mozilla Firefox
Datei Bearbeiten Ansicht Gehe Lesezeichen Extras Hilfe
GOsa²
Hauptmenü
Hilfe
Abmelden
Angemeldet: cajus
Mein Konto
Allgemein
UNIX
Umgebung
Mail
Samba
Konnektivität
Fax
Telefon
Passwort
Administration
Benutzer
Gruppen
Objektgruppen
Abteilungen
Anwendungen
Systeme
FAI
Fax-Sperrlisten
Telefon-Makros
Telefon-Konferenzen
Zusätzliches
Adressbuch
Fax-Berichte
Telefon-Berichte
Systemprotokolle
LDAP-Manager
Automatische Installation
cn=FAIBASE,ou=disk,ou=fai,ou=configs,ou=systems,dc=gonicus,dc=de
Partitionen
Gerät
Name * disk1
Beschreibung Test
Partitions-Einträge
Typ
Dateisystem
Mount-Punkt
Größe in MB
Mount-Optionen
Dateisystem-Option
Bewahren
primary ext3 / 2048-4096 -j Entfernen
primary swap swap 512 Entfernen
logical ext2 Entfernen
Partition hinzufügen
Sichern
Abbrechen

faimond-gui


Faimond-gui
hostname
confdir
defclass
partition
extrbase
debconf
instsoft
configure
tests
savelog
faiend
reboot
demohost
atom03
atom02
atom01
gnomehost

Fakten

- Mehr als 250 detailierte Berichte von Benutzern
- FAI läuft auf i386, amd64, IA64, SPARC, PowerPC, ALPHA
- Ubuntu, CentOS, Suse, Mandrake, ...
- 12k Zeilen Source code (ohne Dokumentation)
- Beispiel Konfiguration ca. 1500 Zeilen

Zusammenfassung

- Homepage: `http://fai-project.org`
- Wiki: `http://wiki.fai-project.org`
- Zwei Maillinglisten, IRC Channel
- CD Images für i386 und amd64
- Regelmäßige Entwicklertreffen
- 11+ Jahre FAI, Erfahrung, Rückmeldungen, Patches durch Benutzer
- Kommerzieller Support: z.B. `fai-cluster.de`

Plane deine Installation

und FAI installiert

deinen Plan!