



UPPSALA
UNIVERSITET

FTP.SUNET.SE 2001 a.p.C.n.



i samarbete med



UPPSALA
UNIVERSITET

1. Sammanfattning

2. Användning, statistik

3. Spegling

4. Utveckling

5. Funderingar

6. Tid & pengar

7. Framtiden

A. Tillgänglighet

B. Spegellista

C. Utrustning

D. Personal

E. Referensgruppen

1. Sammanfattning

ftp.sunet.se har under året skickat ut nära nog dubbelt så mycket som året innan. Lagringskapaciteten har mer än fördubblats, utmatningskapaciteten likaså. Backupkapaciteten har anpassats för detta. Kostnaden för underhåll har trimmats ned cirka 25%. Den underliggande diskstrukturen döljs för användarna, som i stället bara upplever en stor filarea. Söktjänsterna har utvecklats. Antalet speglar var ungefär konstant.

2. Användning, statistik

Det högsta 5-minutersmedelvärdet för utgående trafik som uppmätts under året är 429.6Mbps (23 oktober efter att Red Hat och Mandrake släppt nya versioner). Vi har sett en ökning av trafiken under året med kanske 50%. I bytes räknat är naturligtvis Linux-distributionerna, med Mandrake och Red Hat i spetsen störst, men sett till antalet nedladdningar har program som IRC-klient, Audiogalaxy och winzip legat i topp.

Antal	Filnamn
1491196	/audio/mp3searchtools/AGSetup0606.exe
1247940	/audio/mp3searchtools/CNET-audiogalaxy0605.exe
1178313	/utilities/winzip80.exe
772227	/internet/chat/icq/icq2000b.exe
438738	/internet/communications/HLClientPC1.8.5_Installer.exe
353211	/gamesdomain/shareware/Hacker1_1.zip
297592	/gamesdomain/multiplayer/mods/csv11full.exe
292438	/utilities/zonalarm.exe
289637	/redhat/linux/7.1/en/iso/i386/seawolf-i386-disc1.iso
269347	/redhat/linux/7.1/en/iso/i386/seawolf-i386-disc2.iso
215029	/cnet/win95/audio/mp3searchtools/LimeWireWin.exe
164626	/gamesdomain/demos/3dsrdemo.zip
161171	/cnet/win95/games/adventure-rpg/DiabloIIDemo.exe
146313	/gamesdomain/demos/tdr_all.zip
145956	/distributions/slackware/slackware-8.0/isos/install.iso
142835	/idgames2/planetquake/classicgaming/mame32qa/mame32b-036.zip
122656	/gamesdomain/demos/MaxPayneDemo.exe
116370	/idgames/idstuff/quake/quake106.zip
112362	/gamesdomain/demos/FIFA2001v3.exe
107707	/usr/lib/libc.so.1
105546	/redhat/linux/7.2/en/iso/i386/enigma-i386-disc2.iso
103513	/redhat/linux/7.2/en/iso/i386/enigma-i386-disc1.iso
99614	/vendor/adobe/adobe/photoshop/win/6.x/AdobePS6Tryout.exe
99208	/cnet/win95/games/arcade/zc184.zip
97450	/distributions/mandrake/iso/Mandrake80-ext.iso

Ovan har vi listat de filer som har laddats ner flest gånger under året.

Sökvägarna har kortats ner. För en mer komplett listning Se:

http://ftp.sunet.se/statistik/new/2001/all_year--summed_and_sorted

Fördelningen mellan olika regioner tyder på att runt 20-25% går till Sverige, 10% till övriga Norden, och 20-25% till USA. En del IP-nummer kan vi inte resolvera. Vi räknar med att dessa inte påverkar fördelningen i någon större utsträckning.

Totalt levererade ftp.sunet.se omkring 0.37PB under året. Eftersom vi inte för

statistik på volymen för inhämtning av data för att hålla speglar aktuella är det svårt att säga vilken utväxling detta innebär. Sunet debiterar i dagsläget inte internationell trafik, men om man antar att produktionskostnaden motsvarar 2p/MB som i Storbritannien innebär de 0,1PB som hämtades till Sverige en besparing på MSEK25, vilket står i paritet med vad man kan spara på portokostnader för Tal- och punktskriftsbiblioteket när deras nätdistribution byggs ut.

3. Spegling

Arkivet har i princip inget originalinnehåll utöver bildarkivet som SLU byggde upp, utan allt speglas ifrån andra källor. Vid årets början hade vi cirka 450 speglar. Under året har ett antal speglar tillkommit, bl a:

Progeny Linux
Redmond Linux
Rootlinux
Peanut Linux
Yellowdog Linux
Mandrake Linux för PPC och Sparc
Ultrapenguin Linux
OpenOffice
Xmms
Ximian
ImageMagick
cvsgui

Det har också varit ett antal som har försvunnit under året. Bl a:

GnomeWWW spegeln.
netlib-benchmarks (filerna finns kvar på arkivet)
hartstone (filerna finns kvar på arkivet)
runeberg (avslutad 2000)
The Oxford Text Archive (avslutad 2000)
ConceptBase (avslutad 2002, filer finns kvar)
Directory indexing tool (filer finns kvar)
java-javasoftware (avslutad 2001, filer finns kvar)
macwais (avslutad 2001)
stallman (troligen redan 2000)
mpeg-archive (2001, filerna finns kvar)
microtoft (NetBus Detective) (Avslutad 2001, filerna finns kvar)
uncgi (Avslutad 2002, filerna finns kvar)

Särskilt kan nämnas att vi hela hösten försökt spegla SourceForge, som är hemvist för mycken intressant frivara idag. Tyvärr är konnektiviteten till deras server sådan att vi aldrig lyckats få till stånd en hel spegel.

4. Utveckling

Under året har ett antal utvecklingsprojekt påbörjats (och i vissa fall färdigställts):

- Posthanteringen för arkivet, förutom själva ftpmail-tjänsten, har flyttats in på UU:s centrala postkontor, där virusfiltrering m.m. tillhandahålls. Detta innebär också att man kan nå archive@ftp.sunet.se även om arkivet är otillgängligt.
- FTPsearch är nu åter i drift.
- Mongoose har börjat användas för fritextsökning på arkivet, efter det att UltraSEEK slagit i licenstaket.
- ftp0, ftp1, ftp2, och ftp3 har alla uppgraderats till Solaris 8. ftp4 ligger kvar på Solaris 2.6.
- Genom omskrivningar av ls, locate, m.fl. program har vi kunnat dölja den underliggande uppdelingen i nära dussinet filsystem för arkivet. Tidigare såg användarna detta i form av /pub1 /pub2 o.s.v., vilket kunde verka förvirrande. Numera finns endast /pub vilket ger ett sobrare intryck.
- Från att ha varit begränsad till 2x100Mbps (i praktiken gick vi ut med 198Mbps som 5-minutersmedelvärde i topparna) har utrustningen framför arkivet in till frontmaskinerna ftp1&2, samt en av bakändorna, uppgraderats till 2x1Gbps. Med fortsatt tuning och förbättring av programvara torde systemet under gynnsamma trafikmönster kunna nå upp i 500Mbps ut.
- Som experimentell tjänst har rpmfind tagits upp.
- Datalagringskapaciteten har byggts ut från cirka 900GB till cirka 2.5TB. Av detta har TPB:s del ökat från 70GB till 500 GB.
- Arkivpersonalen har knutit kontakter med andra arkiv runt om i världen, och utbytt tankar och information med en del av dessa.
- Användare har föreslagit ett antal nya speglar. Många har varit mindre Linux-distributioner eller spelorienterade, och i flertalet sådana fall har vi avstått från att ta upp dem. Linux-sektionen har dock utökats. Främst genom att ett antal PPC-distributioner har tillkommit.
- Trafikstatistiken har utökats under året. Förutom veckorapporterna <http://ftp.sunet.se/statistik/> finns även trafiksiffror på utgående trafik på <http://netstat.uu.se/Stat/ftp/router> Veckorapporterna resolveras nu även i högre grad än tidigare. Vilka de populäraste filerna varit kan man se på <http://ftp.sunet.se/statistik/new/> Man bör observera att dessa topp-listor är endast baseras på ftptrafik, och alltså inte inkluderar http och rsync.

5. Funderingar

Under året har vi stött på problem med bl a:

- Ända sedan vi började driva arkivet har Solaris ofta sagt 'Too Many Open Files,' vilket varit förbryllande. Suns tekniker har inte heller kunnat finna orsaken. Situationen verkar ha förbättrats efter uppgradering till Solaris 8. Vi lutar åt att wu-ftp'd tappar öppna filhandtag, och att detta leder till att för många filer hålls öppna.
- Att Deadlock i cacheFS i Solaris kändes besvärande innan vi äntligen fick en crashdump som visade på detta. Att vissa Solaris 7-versioner inte kunde skapa dumpar försvårade felsökningen.
- Att TT5400-arrayen segade plötsligt ner sig under november, och arbetar frenetiskt med ett RAID-set. TransTec:s organisationsförändringar bidrar inte till att få hjälp i felsökningen.
- Att Solaris får slut på minne utan att någon process använder det är inte lätt att diagnosticera. Sun menar på att frontändorna köar upp anrop mot ftp4, vilket får denna konsekvens, varför de förordar uppgradering till GbE även mot backend.
- Att Cisco-switchar dyker när man kaskadkopplar dem var inget vi förutsåg. Nyare IOS ordnar detta.
- Att TPB:s brandmur bygger om paket som passerar den gav länge upphov till mystiska felbeteenden.
- En bug i Apache/PHP gör att barnprocesser inte hanteras korrekt när man slår i taket på antal barn, vilket leder till hängade processer. Problemet verkar allmänt känt, och på sina håll har man ändrat i källkoden för att komma runt det. Trots idogt letande har vi inte funnit någon mera officiell patch, utan har fått hanka oss fram med mer eller mindre väl fungerande work-arounds. Att cacheFS samtidigt gav upphov till deadlocks försvårade undersökning.

6. Tid & pengar

SUNET bekostar underhåll av hårdvara, vilket under året har minskat från kSEK 42/ månad till kSEK30/månad, avskrivning av hårdvara till för oss okänt belopp, och investeringar i tilläggshårdvara. Den senare posten utgörs av cirka kSEK 650 för 2TB disk med tillhörande kringutrustning, cirka kSEK 150 för uppgradering till Gigabit Ethernet på arkivets framsida, och cirka kSEK 17 för cachediskar.

IT-stöd tillställer två heltidstjänster till arkivets förfogande mot en ersättning av kSEK 1438. Nödvändig lokalförhyrning uppgår till kSEK 80. Sedan tillkommer backup kostnad kSEK 150/år. Avstämning av nedlagd tid vid årets slut visar på att IT-stöd nedlagt 2,5 tjänster för driften.

Till detta kommer transmissionslinje och viss nätrelaterad hårdvara, primärt en GSR12000.

7. Framtiden

Nuvarande arkivgeneration härrör primärt ifrån 1998, med diverse uppgraderingar under resans gång. Tanken är att under hösten 2002 byta till en i all väsentlighet ny maskinpark. Ambitionsnivån i diskussionerna är satt till 2Gbps sustained, och 8-10TB disk. Viss fortsatt utveckling av det nuvarande arkivet måste dock ske fram till dess.

Sedan 1999 har Tal- och Punktskriftsarkivet huserat i arkivet, under det att de bedrivit testverksamhet med elektronisk distribution av talböcker. När den verksamheten nu utvidgas kommer TPB att flytta till en egen maskin, vilket underlättar för egna anpassningar. Även fortsättningsvis står Sunet för transmissionskapacitet gemensam för ftp-arkivet.

Man kan förvänta sig en fortsatt ökning av speglingar, föreslagna av användare, driftspersonalen, och referensgruppen.

En gång i världen räckte det att samla på sig ett antal speglar och tillhandahålla dem via ftp för att bli känd. Idag finns det en uppsjö ftp-arkiv av varierande storlek. För att även fortsättningsvis vara i ledningen kan man inte nöja sig med att följa vad alla andra gör, utan behöver flytta fram positionerna i utbudet. I samarbete med referensgruppen letar vi därför nya tjänster att utveckla arkivet med.

Vi har ordnat en licens av SimpliCity, som kan matcha bilder mot varandra och ge ett mått på hur lika de är, för att kunna söka i bildarkivet.

Förbättrad programvara i form av sendfile()-stöd skulle kunna förbättra kapaciteten. Även en ftp-daemon som inte startar en process per koppel, t.ex. ncftpd, skulle kunna medge ökad trafik med nuvarande minnesmängd i frontändorna.

Vi skulle vilja utveckla kontakterna med våra användare.

Möjligheten att, troligen i samarbete med lämpligt företag, erbjuda distribution via CD har diskuterats men inte närmare studerats.

Man skulle kunna tänka sig att viruskolla material på arkivet.

Arkivet har för det kommande året gjort följande budgetäskanden:

1. Nytt arkiv
2. Oförutsedda kostnader
3. Kort, Quad-Ethernet
4. ncftpd-licenser
5. Profilmaterial
6. Diskar.
7. Cache-diskar
8. GbE-kort (lånade kort används under december)
9. Lab-barandmur
10. NORDUnet-konferensen
11. VxVM-licens
12. Ny disk-IO-arkitektur
13. Switch (kan strykas om man tar dyrare kort som under låneperioden)
14. Eget gosedjur som symbol för ftp.sunet.se
15. Sökmotor (kan nog strykas, men det var oklart vid budgeteringen)

A. Tillgänglighet

Arkivet kan från tid till annan vara otillgängligt, helt eller delvis, av många olika orsaker. Några typer av avbrott kan dock identifieras:

Nätproblem mot Stockholm	4 ggr	stillestånd.
Nätproblem lokalt	2 ggr	stillestånd.
Too Many Open Files	flertal ggr	begränsad framkomlighet.
CacheFS deadlock i Solaris	flertal ggr	begränsad framkomlighet.
Automounter i Solaris	flertal ggr	begränsad framkomlighet.
Uppgraderingar	5 ggr	stillestånd.
Hårdvaruproblem	2 ggr	begränsad framkomlighet.
GbE uppgradering	4 ggr	stillestånd.

Experiment med automatisk loggning av nertid inifrån systemet har stött på problem i och med att TCP-koppel har långa timeouter, varför upp till 40 minuter missas av avbrott.

B. Spegellista

X
XR6.1
XFree86
XR6.3
XR5-contrib
XR6-contrib
ezd
aui
fvwm
xv
aburto-benchmarks
wpi-benchmarks
007
netlib-benchmarks
musbus
NAS
SSBA
hartstone
presno
gutenberg
libellus
wiretap-classics
Finnish
lbl-conferencing
ivs
cvsgui
orionserver
bsd-db
lincks
DiamondBase
mysql
mood
exodus
shore
OBST
texaspstore
texaspstore-paper
ConceptBase
coral
GRAS
RDB
smart
siod
tsql
refdbms
wu-ftp
funic-ftp
ncftp
ncftp
libncftp
net2+mea-ftp
llnlxftp
llnlxdir
ftp-archive

ftpmail
squirell
ftpcat
ftpindex
mirror
fmirror
ftpd-ls
mirrorftp
iafa
x-plane
gamesdomain
pc-games-idgames
pc-games-idgames2
3drealms
flight-sim
planetquake
Angband
nethack
life
xbl
xconq
xconq7
crossfire
egcs
gimp
gimp-manual
SNePS
ispell
gnuplot
elisp-archive
OSullivan-elisp
easy-edit
vm
lpf
gnus
gopher
gopherpup
gopherapp
MacGopher
pnlinfo-mac
gn
bbgopher
Collage
wscrawl
xmk
DistEdit
GroupKit
rfc
internet-drafts
ietf
iesg
isoc
yanoff
jargon
nethlp
inetdirs

matrix
pride
cni-net-guides
isode
PC-scheme
wxwin
intellij
compilers-list
Lisp-Utilities
tcl
expect
smalltalk
python
hytelnet
silverplatter
YDL
linux-funet
debian-non-US
peanut-linux
mklinux-apple
debian
linux-alsa
mklinux-pre-R1
mklinux-pre-R1-2
mandrake-crypto
bestlinux
swe-doc
info-mac
info-mac/rec
dartmouth-Mac
newswatcher
MPEGencoder
nih-image
mailstrom
NCSA-Telnet-Mac
NCSA-Telnet-Mac-svenska
Mac-Collage
macweb
simple-http-editor
cap
netatalk
alpha
HTML_Editor
postfix
sendmail
emil
minuet
NUPop
POPmail-PC
RFDmail
eudora
pegasus
tkrat
imap
pine
elm

mh
exmh
majordomo
listserv
metamail
mpack
imdb
spacebattles
kidlink
3dwm
Berkeley-multimedia
mpeg-netcom
mplex
jpeg
gd
png
AF
gsm
HDF
pov-ray
xmms
imagemagick
geomview
cma
cma-files
CMU-SNMP
ut-snmv2
tcpdump
ttt
nikhef-network-tools
NNStat
icmpinfo
fping
nettest
netperf
ttcp
nocol
netmon
arpmon
tcpview
xtacacsd
ipxping
OSItrace
ip-routing
zebra
merit-research
isc
dns
ntp
cslip
SVR4slip
ppp
ethernet-codes
tkined
irc
news-uunet

cnews
rkive
tin
hampsons-gopher
wsarchie
NIDR
Hyperwave
VRweb
essence
x500-ldap
mirror-brunel
netfind
willow
winwillow
z3950-loc
oclc-z39.50_client_api
BER_utilities
simberascii
harvest-uk
sfinger
icsi-finger
BeOS
soss
NCSA-Telnet-PC
trumpet
pc-nfs-york
os2-hobbes
wintalk
qvt-net
qws3270
ewan
squiggle
sticky
pmlpr
timesync
WinVN
NT-emwac
NONAGS
freeware95
msdos-archives
win3
win95
winnt
win2000
winsock-stardust
winsock
winsock-indstate
winSqueeze
ntemacs-docs
gnu-win32
Dr-Fun
ACiD
12hr-ISBN-JPEG
vendor-microsoft
vendor-microsoft-developer
bio-indiana

gnupg
security-cert-de
security
sourceforge
freetype
sgml
cost
RTF
wp2x
TeX
Lout
DOSgroff
behandler
Quikscript
ghostscript
rsync
am-utils
most
lsof
OpenBSD
FreeBSD
usenet-by-hierarchy
usenet-by-group
usenet-sources
vendor-sco
skunkware
adobe
vendor-sun-doc
swish-e
wais-cornell-hacks
wais-iubio
wais-perl-server
whois
winwhois
schoultz-whois++
hamilton-whois++
WMosaic-svenska
CU_html
roxen
www
xmosaic
netscape
netscape-communicator-english-479
mozilla
lynx
DosLynx
WebEdit
HTMLed
chimera
Squid
marmot
cl-http
warmlist
webcopy
wget
ffw

rtftohtml
fm2html
getstats
web500gw
miftran
html-2.0
html-helper
www-fielding
SFgate
ForumNews
simon
wusage
mapedit
www-tools_uni-karlsruhe
treelink
htmltext
gwstat
mapmarker
dtd2html
uncgi
ASHE
asWedit
hyperactive
analog
polymap
12hr
Imagemagick
allmacintosh
cma
netsaint
presno
simtel
xmm
CPAN
LDP
apache-dist
apache
asplinux
debian-iso
debian-iso_temp
delphi
ethereal-ftp
ethereal-web
gnomeftp
gnu
gnuftp
gnupg
helixwww
hwb
iozone
kde
kernel-crypto
kernel-linux
kernel-software
linuxppc
mandrake-local

mandrake-new
mysql
mysql_sourceforge
netbsd
opera
postgresql-www
postgresql
redhat-local
redhat-new
redmondlinux
rootlinux
rpmfind
samba
scene-org
simtelnet
slackware
sourceforge
sunfreeware
suse
xemacs
ximian
aminet
kde
simtel.net
tucows
TolkiensOxford.mirror
webcopy
back-forward
curl
foldoc
geda-ftp
geda-web.2
geda-web
geda
gnu
html_dict
isc
ivalley
kerberos
linux_cmu
lrrd
magnicomp
math-t
mrtg
netsaint
no_sugar
rrdtool
warriors
wnserver
www.cipav.org.co
wxwin

C. Utrustning

Vid årets slut bestod arkivet i huvudsak av följande hårdvara:

GSR12000 (router) för anslutning mot Stockholm.

Cisco 3524 & 3508 (switchar).

BiFrostar, två stycken (GbE-utförande).

Sun E450, två stycken som svarar på ftp.sunet.se

Sun 2200 för datainsamling

Sun 2100 för test och intern konnektivitet

Sun E4500 för att hantera I/O-systemet och sökmotorer.

Sun F3500 diskarray.

TransTec 5400 diskarray.

PC som console-server.

IT-stöds backupsystem, bestående av en NSM med tillhörande 3584-robot, används även för ftp-arkivet.

IT-stöd tillhandahåller även syslog-funktion för arkivet.

Programvaran utgörs huvudsakligen av fri programvara. De viktigaste är wu-ftpd, apache, rsync, wget, locate, mnogosearch och mirror.pl.

D. Personal

Den dagliga driften sköts primärt av Andreas Wallin och Emil Pedersen. Lars-Owe Ivarsson och Dan Pettersson är också inblandade i arkivdriften. På nätsidan arbetar Hans Wassén och Robert Olsson. För upphandlingsstöd svarar Ragnar Ahlberg. Många andra på IT-stöd bidrar ifrån tid till annan beroende på kompetens som behöver tillföras arkivet.

E. Referensgruppen

Till arkivet är knuten en referensgrupp som bistår med vägledning avseende framförallt arkivets innehåll, men även andra frågor som berör arkivet i stort. Gruppen kan nås på adress referens@ftp.sunet.se, vilken även går till driftspersonalen, och bestod under 2001 av följande personer:

Mattias Pantzare, Luleå

Börje Josefsson, Luleå

Joakim Björklund, Linköping

Torbjörn Olofsson, Dalarna

Dennis Sjögren, Dalarna

Peter Byström, Örebro

Magnus Taube, Halmstad

Johan Hofvander, Anoto AB (förutvarande arkivarie)

