

# **Andrew File System / OpenAFS**

**Ein globales, verteiltes Dateisystem**

Seminarvortrag "Dateisysteme" SS07 Uni Heidelberg

10. Juli 2007

---

# Übersicht

- Merkmale
- Geschichte
- Kurzeinschub Kerberos
- Aufbau
- Hard Facts
- Einsatzbeispiel
- Live Demo
- Ausblick

# Gliederung

- **Merkmale**
- Geschichte
- Kurzeinschub Kerberos
- Aufbau
- Hard Facts
- Einsatzbeispiel
- Live Demo
- Ausblick

## Merkmale

- **verteilt**

- Aufgaben beliebig auf Rechner verteilt

- **global**

- Rechner beliebig im Netz verteilt
- Globaler Namensraum

- **sicher**

- Traffic verschlüsselt
- Authentifizierung über Kerberos

- **robust**

- Integrierte Replikation
- Integrierte FailOver-Fähigkeit
- Integriertes Backup-System

## Merkmale: Unterstützte Betriebssysteme

- Linux2.4: Server+Client
- Linux2.6: Server+Client
- Windows: Client, Experimentelle Server ab W2k
- OS X: Server+Client
- Solaris: Server+Client ab 2.0
- AIX5\*: Server+Client
- KEIN BSD!

# Gliederung

- Merkmale
- **Geschichte**
- Kurzeinschub Kerberos
- Aufbau
- Hard Facts
- Einsatzbeispiel
- Live Demo
- Ausblick

## **Geschichte: Timeline**

### **1984**

- Im Rahmen des Project Andrew an der Carnegie Mellon University entwickelt
- Benannt nach Andrew Carnegie, Gründer des Carnegie Institute of Technology

### **1989**

- Namensgebung
- Transarc gegründet zur kommerziellen Vermarktung
- Zu passiv, kein richtiger Erfolg

### **1998**

- Transarc von IBM übernommen

### **2000**

- OpenAFS 1.0 unter der IBM PL veröffentlicht

## 2007

- Aktuelle Version: 1.4.4 bzw. 1.5.20

## Notes

- 1-Dec-2006 - Announcing OpenAFS "Works with Windows Vista"
- 19-Mar-2006 - OpenAFS 1.4.4 released
- 4-Apr-2007 - OpenAFS 1.5.18 released
- 18-May-2007 - OpenAFS 1.5.20 released

## Geschichte: Implementationen

### Arla

- freie Implementation der KTH zu Transarc-Zeiten
- Nur Client, aber viele OS (z.B. \*BSD)

### MR-AFS

- Basiert auf Transarc-AFS
- Nur Server
- Tertiärspeicheranbindung per HSM

### HostAFS

- Mini-Server
- Einfache Freigabe von Verzeichnissen im AFS
- ohne ACLs

## kAFS

- im Linux-Kernel integriert
- Nur Client
- read-only
- nur für Spezialaufgaben nutzbar

# Gliederung

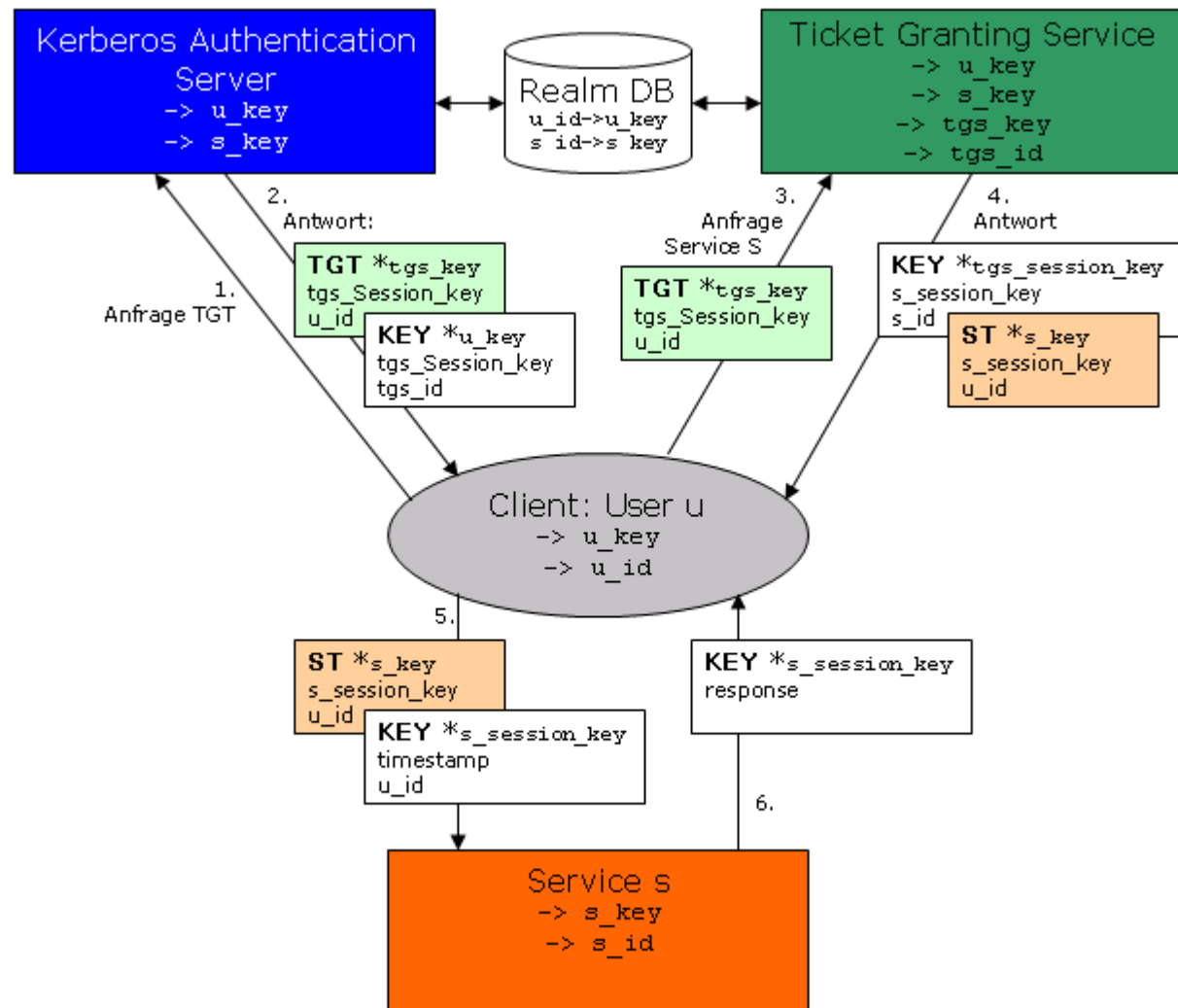
- Merkmale
- Geschichte
- **Kurzeinschub Kerberos**
- Aufbau
- Hard Facts
- Einsatzbeispiel
- Live Demo
- Ausblick

# Kerberos

## Was ist der Höllenhund?

- Verteilter Authentifizierungsdienst
- für offene und unsichere Computernetze
- Single Sign on
- authentisiert sowohl den Server ggü dem Client
- als auch den Client ggü dem Server
- als auch sich selbst ggü beiden
- TGT
- SessionKey-verschlüsselt

# Kerberos: Skizze



# Gliederung

- Merkmale
- Geschichte
- Kurzeinschub Kerberos
- **Aufbau**
- Hard Facts
- Einsatzbeispiel
- Live Demo
- Ausblick

# Aufbau: Terminologie

## Zelle

- Unabhängige Instanz
- Eigener Namespace
- kann verteilt sein
- User haben Heimatzone
- können aber in mehreren Zellen Mitglied sein
- globale Datenbank der Zellen

## Aufbau: Terminologie

### Volume

- Container für Dateien
- unterstützt Quota
- wird mit Mountpoints in Namensraum der Zelle eingehängt

### Partition

- Physikalischer Speicherort der Volumes auf dem Server

### MountPoint

- linkt Volume in Verzeichnis

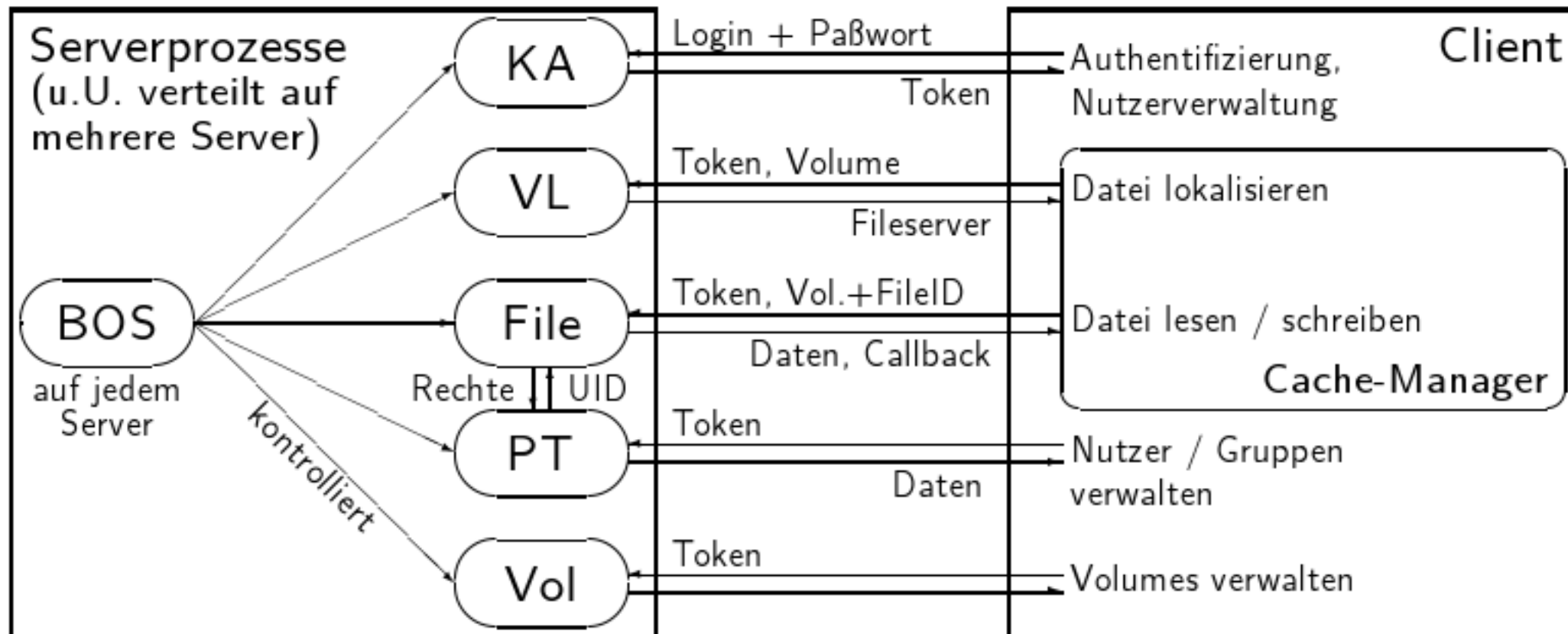
# Aufbau: Konzept

## Ganzheitliches Konzept

- BOS
  - Basic Overseer Server
  - Administrative Instanz
  - läuft auf jedem phys. Server
- KA
  - Kerberos Authentication Server
  - Authentifizierung, Nutzerverwaltung
  - Integrierter Kerberos 4
- VL
  - Volume Location Server
  - Dateiserververwaltung mittels VLDB

- File
  - File Server
  - liefert Daten aus, empfängt Änderungen
- PT
  - Protection Server
  - Verwaltet ACLs und setzt diese ggü File durch
- Vol
  - Volume Server
  - Volumemanagement (Erstellen, Migrieren, löschen)
- Update
  - Distribution von Software- und Konfigurationsupdates
- Backup
  - Koordiniert Backup- und Restore-Operationen
- NTPD

## Aufbau: Skizze





# Aufbau

## auf dem Client (aka. Cache-Manager)

- Lokaler Disk-Cache (!journal)
- mit Callback
- Kernel-Modul + Userland
- PAM-Modul für Kerberos-Tickets und AFS-Token evtl. sinnvoll

## Aufbau: ACLs

- Zugriff auf non-public nur mit AFS-Token (wird aus Kerberos-Ticket generiert)
- Rechteverwaltung über ACLs pro Verzeichnis(!)  
Speziell: system:anyuser system:authuser system:administrators
- ACLs (+-,ug):
  - l lookup
  - r read
  - i insert
  - w write
  - d delete
  - k lock
  - a admin

# Gliederung

- Merkmale
- Geschichte
- Kurzeinschub Kerberos
- Aufbau
- **Hard Facts**
- Einsatzbeispiel
- Live Demo
- Ausblick

## Hard Facts: Limitierungen

- Pro Zelle max. 254 Dateiserver
- Pro Dateiserver max. 255 Datenpartitionen
- Pro Datenpartition max 4 TiB (32Bit \* Blockgröße(1kB))
- Pro Verzeichniss max. 64435 dentries
- Volumenamen max. 22 Zeichen
- Pro Volume max. 2 TiB

## Hard Facts: Diverses

- root hat keine Sonderrechte
- 1 v4-IP pro Server
- !NAT
- mbox vs. maildir

## Hard Facts: Anekdote

If you are not an Internet site, it is best to choose a unique Internet-style name, particularly if you plan to connect to the Internet in the future. AFS Product Support is available for help in selecting an appropriate name. There are a few constraints on AFS cell names:

It must end in a suffix that indicates the type of institution it is, or the country in which it is situated. The following are some of the standard suffixes:

- \* .com For businesses and other commercial organizations. Example: abc.com
- \* .edu For educational institutions such as universities. Example: stateu.edu
- \* .gov For United States government institutions.
- \* .mil For United States military installations.

Other suffixes are available if none of these are appropriate. You can learn about suffixes by calling the Defense Data Network [Internet] Network Information Center in the United States at (800) 235-3155. The NIC can also provide you with the forms necessary for registering your cell name as an Internet domain name. Registering your name prevents another Internet site from adopting the name later.

## Hard Facts: Eindrücke

# ls /afs/

1ts.org engr.wisc.edu lcp.nrl.navy.mil rhic.bnl.gov

acm.uiuc.edu eng.utah.edu le.infn.it rl.ac.uk

ams.cern.ch epfl.ch Inf.infn.it rose-hulman.edu

andrew.cmu.edu es.net lngs.infn.it rpi.edu

anl.gov ethz.ch lrz-muenchen.de rrz.uni-koeln.de

asu.edu extundo.com isa.umich.edu rz.uni-jena.de

athena.mit.edu f9.ijs.si math.unifi.it sanchin.se

atlass01.physik.uni-bonn.de fnal.gov md.kth.se sbp.ri.cmu.edu

atlas.umich.edu fusione.it mech.kth.se scoobydoo.psc.edu

azubi.lan glue.umd.edu mekinok.com scotch.ece.cmu.edu

ba.infn.it gppc.de membrain.com s-et.aau.dk

bazquux.org grand.central.org meteo.uni-koeln.de setfilepointer.com

biocenter.helsinki.fi hackish.org midnightlinux.com sinenomine.net

bme.hu hallf.kth.se mpe.mpg.de sipb.mit.edu  
caspur.it hep.caltech.edu msc.cornell.edu slackers.net  
cats.ucsc.edu hephy.at msu.edu slac.stanford.edu  
cede.psu.edu hep.man.ac.uk nada.kth.se soap.mit.edu  
cern.ch hep.sc.edu ncsa.uiuc.edu sodre.cx  
chem.cmu.edu hep.wisc.edu nd.edu stacken.kth.se  
ciemat.es i1.informatik.rwth-aachen.de nersc.gov su.se  
citi.umich.edu iastate.edu net.mit.edu syd.kth.se  
clarkson.edu ic-afs.arc.nasa.gov nikhef.nl tgrid.it  
club.cc.cmu.edu icemb.it nimlabs.org tproa.net  
cmf.nrl.navy.mil ictp.trieste.it nomh.org tu-bs.de  
coed.org idahofuturetruck.org northstar.dartmouth.edu tu-chemnitz.de  
cs.cmu.edu ies.auc.dk oc7.org umbc.edu  
cs.pitt.edu ifca.unican.es openafs.org umich.edu  
cs.rose-hulman.edu ifh.de pdc.kth.se umr.edu  
cs.stanford.edu ific.uv.es phy.bris.ac.uk uncc.edu  
cs.uwm.edu in2p3.fr physics.unc.edu uni-bonn.de

cs.wisc.edu infn.it physics.wisc.edu uni-freiburg.de  
dapnia.saclay cea.fr ing.uniroma1.it physik.uni-freiburg.de uni-hohenheim.de  
dbic.dartmouth.edu ipp-garching.mpg.de physik.uni-mainz.de uni-mannheim.de  
dementia.org ir.stanford.edu physik.uni-wuppertal.de uni-paderborn.de  
desy.de isk.kth.se physto.se urz.uni-heidelberg.de  
dev.mit.edu italia pi.infn.it usatlas.bnl.gov  
dia.uniroma3.it it.kth.se pitt.edu vn.uniroma3.it  
e18.ph.tum.de itp.tugraz.at p-ng.si wam.umd.edu  
ece.cmu.edu jpl.nasa.gov psc.edu wu-wien.ac.at  
eecs.harvard.edu kfki.hu psi.ch  
e.kth.se kloe.infn.it psm.it  
enea.it laroia.net qatar.cmu.edu

## Eindrücke

```
#ls -l /afs/urz.uni-heidelberg.de/usr/pci/c62/WWW/  
total 2262  
drwxr-xr-x 2 2090 202 2048 2000-11-14 12:04 Files  
-rw-r— 1 2090 202 59386 2000-11-15 13:18 gz.cdr  
-rw-r-r- 1 2090 202 571 2000-11-15 13:22 index.html  
-rw-r-r- 1 2090 202 397 2000-02-24 12:03 java1.html  
-rw-r-r- 1 2090 202 14456 2000-11-13 12:02 PCI_F_Seminar.htm  
-rw-r— 1 2090 202 2230186 2000-11-14 12:03 PCIF.zip  
-rw-r-r- 1 2090 202 622 2000-02-24 11:58 TA.class  
-rw-r-r- 1 2090 202 394 2000-02-24 11:58 TA.java  
-rw-r— 1 2090 202 3082 2000-11-15 13:23 test1.htm  
-rw-r-r- 1 2090 202 403 2000-02-24 11:56 TestAW1.java
```

# Gliederung

- Merkmale
- Geschichte
- Kurzeinschub Kerberos
- Aufbau
- Hard Facts
- **Einsatzbeispiel**
- Live Demo
- Ausblick

# Einsatzbeispiel

## InstantAFS

- "Reinschütten, umrühren, AFS!"
- Sammlung von Perl-Scripten und Konfigurationen
- Debian-Paket für die schnelle Einrichtung einer AFS-Zelle
- sehr gute Dokumentation
- Demozelle

## Einsatzbeispiel: rivers.de

### Das Forschungsinstitut rivers.de besteht aus 2 Teilen:

- Einem grossen Hauptgebäude
- Einer weit entfernten Aussenstelle

jeweils 1 geswitchtes Netz  
verbunden über 2Mbit-I-Net  
3 TiB in Hauptgebäude  
1 TiB in Aussenstelle

## Einsatzbeispiel: rivers.de

### Warum nicht 2 Zellen?

- Nachteile:

- Die Benutzerdatenbanken (Kerberos5, PTDB) müssen abgeglichen werden.

- Benutzer bräuchten mehrere Tokens

- AFS hält keine Bordmittel bereit, um Backups zellenübergreifend zu organisieren

- Vorteile:

- Die Zellen sind unabhängig (solange auch das DNS unabhängig ist)

## Einsatzbeispiel: rivers.de

### Setup Hauptgebäude (10.1.0.0/255.255.0.0):

- Serverraum1
  - orinoco Funktion: (primärer) Datenbankserver IP: 10.1.1.10
  - mekong Funktion: Fileserver IP: 10.1.2.1
- Serverraum2
  - rio-grande Funktionen: Fileserver, Datenbankserver IP: 10.1.1.20
  - ganges Funktion: primärer Backupserver (Portoffset 0) IP: 10.1.2.2

## Einsatzbeispiel: rivers.de

### Setup Aussenstelle (10.2.0.0/255.255.0.0):

- Serverraum

rubicon Funktionen: Datenbankserver, Backupserver ( butc mit Portoffset 1) IP:  
10.2.1.10

volga Funktion: Fileserver IP: 10.2.2.1

yang-Tze Fileserver IP: 10.2.2.2

## Einsatzbeispiel: rivers.de

### Fileserver

- sollten immer üppig dimensioniert sein.
- ein 2. Prozessor bringt ordentlich Gewinn.

### Datenbankserver

- dürfen Sparrechner sein
- ohne viel RAM, HDD und CPU.

### Backupserver

- brauchen für die Kompression der Backups viel Rechenleistung

- Arbeitsspeicher können sie nie genug haben, da dann die temporär unkomprimiert auf Platte liegenden Tape-Images im Cache bleiben
- RAID's werden benutzt, um komprimierte Tape-Images zu lagern.

## Einsatzbeispiel: rivers.de

### Stärken des Setups:

- Die Datenhaltung soweit wie möglich dezentral
- IP-Adressen geschickt gewählt
  - Datenbankserver niedrige
  - genügend Platz zwischen DB-Server
- geswitcht wo nicht geroutet werden muss
- Rolle des Backup-Servers der Aussenstelle könnte auch einer der Fileserver übernehmen

## Einsatzbeispiel: rivers.de

### Schwächen des Beispiels

- Datenbankserver sollten dedizierte Rechner sein.
- Das Backup wird auf ganges ausgelöst und überwacht. Fällt die Verbindung zwischen den Institutsteilen aus, wird in der Aussenstelle kein Backup durchgeführt.
- Ein Backupserver und ein Fileserver ( ganges und rio-grande ) stehen im selben Raum

# Gliederung

- Merkmale
- Geschichte
- Kurzeinschub Kerberos
- Aufbau
- Hard Facts
- Einsatzbeispiel
- **Live Demo**
- Ausblick

# Gliederung

- Merkmale
- Geschichte
- Kurzeinschub Kerberos
- Aufbau
- Hard Facts
- Einsatzbeispiel
- Live Demo
- **Ausblick**

## Ausblick

- *2007Q2* - Rx over TCP
- *2007Q3* - GSSAPI/KRB5 (+ multiple encryption)
- *2008Q1* - IPv6

## Quellenangaben

- <http://b4mad-service.net/datenbrei/archives/2006/03/08/openafs-server>
- [http://de.gentoo-wiki.com/OpenAFS\\_mit\\_MIT-Kerberos5](http://de.gentoo-wiki.com/OpenAFS_mit_MIT-Kerberos5)
- <http://www.gentoo.org/doc/en/openafs.xml>
- <http://sascha.silbe.org/learn/afs/vortrag/global2.2.html>
- [http://de.wikipedia.org/wiki/Andrew\\_File\\_System](http://de.wikipedia.org/wiki/Andrew_File_System)
- <http://fbo.no-ip.org/cgi-bin/twiki/view/Instantafs/WebHome>
- <http://www.openafs.org/>
- <http://fbo.no-ip.org/cgi-bin/twiki/view/Instantafs/DokuMentation>
- <http://netzsure.de>