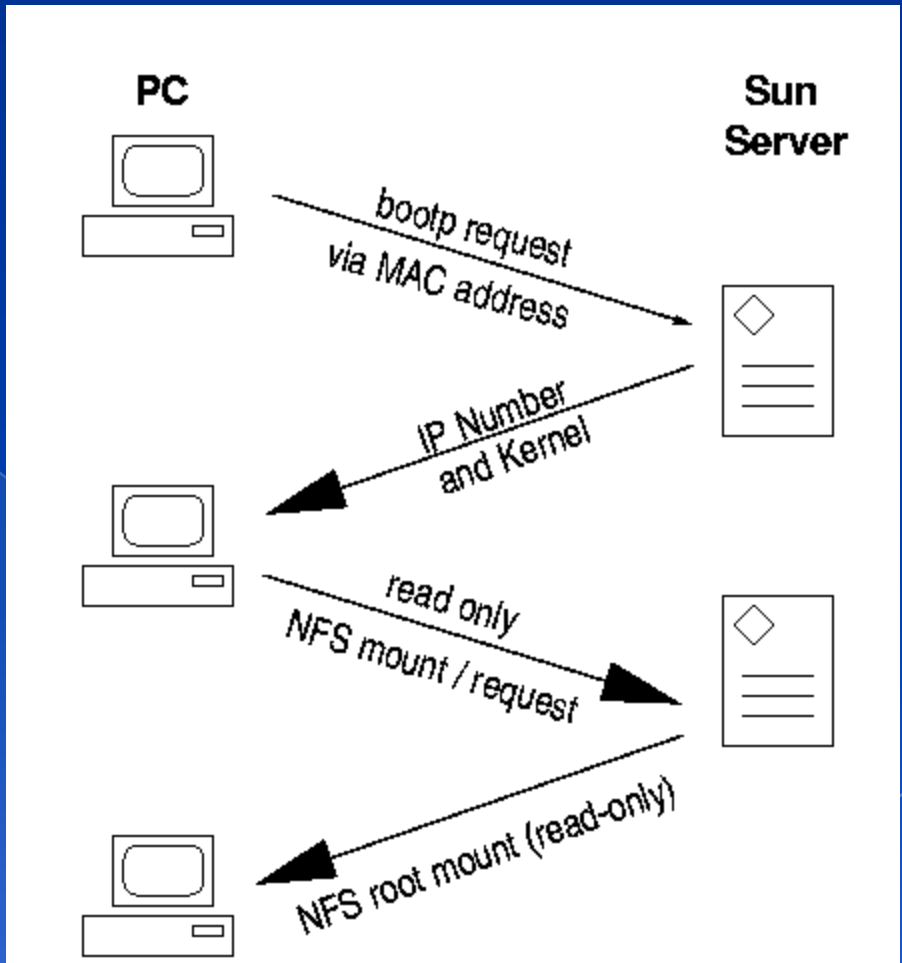
Chapter 17
The Network File System

File System over Network

- Network File System
- Andrew File System
- NetWare Core Protocol
- Server Message Block
- Common Internet File System

NFS

- Share filesystem to other hosts via network
- NFS History
  - Introduced by Sun Microsystems in 1985
  - Originally designed for diskless client-server architecture



Components of NFS

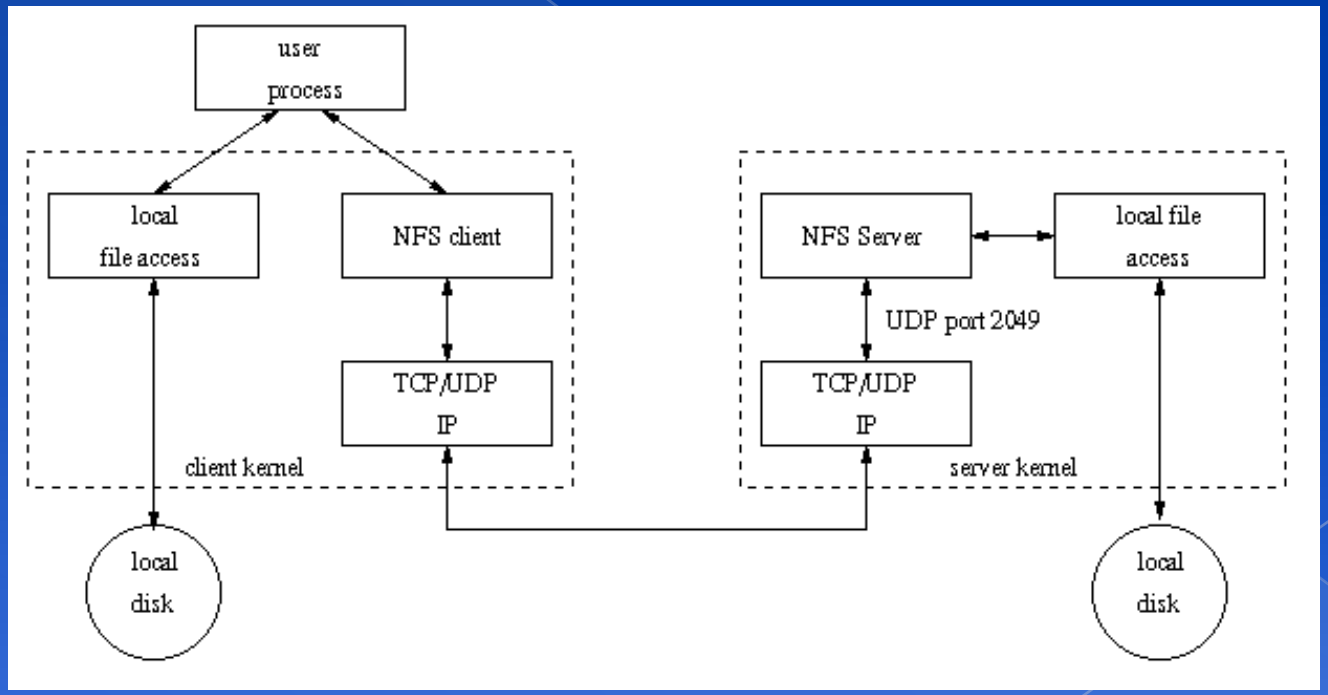
- Including
  - Mounting Protocol
  - Mount Server
  - Daemons that coordinate basic file service
  - Diagnostic utilities

Components of NFS – mounting protocol (1)

- NFS (RFC1094)
- NFSv2 (RFC1094, March 1989)
  - Synchronous write
  - V2 NFS server must commit each modified block to disk before replying to NFS client
  - Cause long delay when there is a NFS write operation
- NFSv3 (RFC1813, June 1995)
  - Asynchronous write
  - Provide increase performance and better support for large files
- NFSv4 (RFC3530, April 2003)
  - Stateful protocol

Components of NFS – mounting protocol (2)

- Sun's ONC distributed computing standards
  - NFS client → RPC → Transport Layer → …
  - Transport Layer
    - UDP: Lack congestion control
    - TCP: become more suitable



Components of NFS – Server-side NFS (1)

- NFS Server
  - Export sharing filesystem
    - System dependent
  - Waiting for "**mount request**"
    - mountd (rpc.mountd) daemon
  - Waiting for "file access request"
    - nfsd (rpc.nfsd) daemon

Components of NFS – Server-side NFS (2)

- Exporting filesystem
  1. Edit export configuration file
     - Each line is "what to export and how"
  2. Reload related daemons

System	Exports info file	How to reload
FreeBSD	/etc/exports	/etc/rc.d/mountd reload
Linux	/etc/exports	/usr/sbin/exportfs -a
Solaris	/etc/dfs/dfstab	/usr/sbin/shareall
SunOS	/etc/exports	/usr/sbin/exportfs -a

Components of NFS – Server-side NFS (FreeBSD.1)

- Exporting filesystem
  - /etc/exports
    - White-space separated
    - Format: directory-list options-list client-list

Option	Description
-ro	Exports read-only, default is (read-write)
-alldirs	Allow any subdirectory to be mounted
-maproot=user	Maps root to the specified user.
-mapall=user	Maps all UIDs to the specified user.

Client	Description
hostname	Host name (ex: mailgate ccserv)
netgroup	NIS netgroups
-network -mask	-network 140.113.235.0 -mask 255.255.255.0

Components of NFS – Server-side NFS (FreeBSD.2)

- Example of /etc/exports

```
/raid    -alldirs –maproot=root csmailgate backup
/raid    -alldirs –maproot=nobody –network 140.113.235 –mask 255.255.255.0
/home    -ro -mapall=nobody –network 140.113.235.0 –mask 255.255.255.0
/usr/src /usr/obj –maproot=0 bsd_cc_cs
```

- Reload daemons
  - # /etc/rc.d/mountd reload

Components of NFS – Server-side NFS (Linux.1)

- Exporting filesystem
  - /etc/exports
    - Format: directory client-list-with-option
    - Ex: /home1 bsd5(ro)

Client	Description
hostname	Host name (ex: mailgate ccserv)
@netgroup	NIS netgroups
ipaddr/mask	CIDR-style specification (ex: 140.113.235.2/24)
Wild cards * ?	FQND with wild cards (ex: bsd*.cs.nctu.edu.tw)

Components of NFS – Server-side NFS (Linux.2)

Option	Description
ro,rw	Read-only, Read-write (default)
rw=list	Hosts in the list can do rw, others ro only
root_squash	Maps UID 0 and GID 0 to the value of anonuid and anongid (default)
no_root_squash	Allow root access
all_squash	Maps all UID and GID to anonymous one
subtree_check	Check that the accessed file is in the appropriate filesystem and in the exported tree.
no_subtree_check	Disables subtree checking
anonuid=xxx	Related to root_squash
anongid=xxx	Related to root_squash
secure	Require remote access from privileged port
insecure	Allow remote access from any port
noaccess	Prevent access to this dir and it's subdir

Components of NFS – Server-side NFS (Linux.3)

- Example of /etc/exports

```
/home1          sun*.cs.nctu.eud.tw(rw)
/home2          @sun_cc_cs(ro)  dragon(rw,no_root_squash)
/home           cspc1(rw,all_squash,anonuid=150,anongid=100)
/ftp/pub        (ro,insecure,all_squash)
/users          *.xor.com(rw)
/users/evi      (noaccess)
```

- Run /usr/sbin/exportfs
  - # /usr/sbin/exportfs –a
    - Maintain /var/lib/nfs/xtab table which is read by mountd

Components of NFS – Server-side NFS (Solaris.1)

- Exporting filesystem
  - /etc/dfs/dfstab
  - Each line will execute “share” command to export one NFS
    - [format] share –F nfs –o option-list directory
    - Ex: share –F nfs –o rw=ccbsd5.csie.nctu.edu.tw /home2
- Run shareall command
  - % /usr/sbin/shareall

Client	Description
hostname	Host name (ex: mailgate ccserv)
netgroup	NIS netgroups
IP networks	@CIDR-style specification (ex: @140.113.235.2/24)
DNS domains	.xxx.yyy any host within the domain (ex: .nctu.edu.tw)

Components of NFS – Server-side NFS (Solaris.2)

Option	Description
ro,rw	Read-only to all, Read-write to all
ro=list, rw=list	Hosts in the list can do ro/rw
root=list	Lists hosts permitted to access this filesystem as root. Otherwise, root access from a client is equivalent to by "nobody"
anon=xxx	Specify the UID to which root is remapped. Default is "nobody"
anongid=xxx	Related to root_squash
nosub	Forbids clients to mount subdirectories
nosuid	Prevents setuid and setgid from being created

Components of NFS – Server-side NFS (3)

- nfsd daemon
  - Handle NFS file access request from NFS clients
  - Number of nfsd is important
    - Too small, some NFS request may be not served
    - Too large, load will be high
- In FreeBSD
  - Specify nfsd options in /etc/rc.conf
    - nfs_server_enable="YES"
    - nfs_server_flags="-u –t –n 4"

nfsd(8), nfsiod(8)

Components of NFS – client-side NFS (1)

- NFS Client
  - Mount NFS filesystem first
  - Access file under NFS filesystem
- mount command
  - [format]
    - *mount* [-o options] *host:directory mount-point*
  - *Ex:*
    - **% mount –t nfs ccbsd4:/home/www /home/nfs/www**
- /etc/fstab (/etc/vfstab in Solaris)
    - **% mount -a -t nfs (FreeBSD, Linux)**
    - **% mount -a -F nfs (Solaris)**

```
# Device            Mountpoint      FStype   Options     Dump  Pass#
dragon:/usr/man     /usr/man        nfs      ro,bg,soft  0     0
ccserv:/spool/mail  /var/mail       nfs      rw,bg,intr  0     0
```

Components of NFS – client-side NFS (2)

- NFS mount flags

mount_nfs(8)

Flag	Systems	Description
ro or rw	S,L,F	Mount the NFS as ro or rw
bg	S,L,F	If failed, keep trying in background
hard	S,L	If server down, access will keep trying until server comes back
soft	S,L,F	If server down, let access fail and return error
intr, nointr	S,L,F	Allow/Disallow user to interrupt blocked access
retrans=n	S,L,F	# of times to repeat a request before error return
timeo=n	S,L,F	Timeout period of requests (tens of seconds)
rsize=n	S,L,F	Set read buffer size to n bytes
wsize=n	S,L,F	Set write buffer size to n bytes
vers=n	S	Selects NFS v2 or v3
nfsv3,nfsv2	F	Selects NFS v2 or v3
proto=prot	S	tcp or udp
tcp	L,F	Select TCP. UDP is default

Components of NFS – client-side NFS (3)

- Client side daemons that enhance performance
  - biod (block I/O daemon, or called nfsiod)
  - Perform read-ahead and write-behind caching

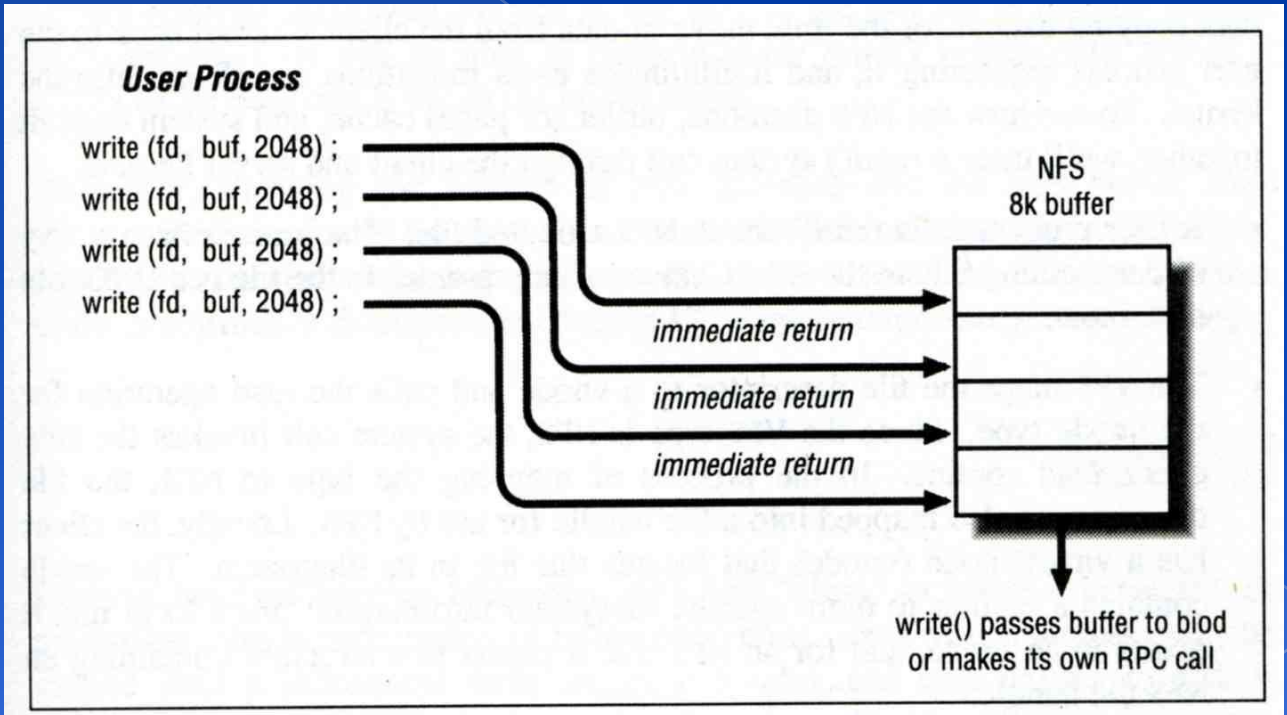


Figure 6-2. NFS buffer writing

Components of NFS – NFS Utilities (1)

- nfsstat
  - Display NFS statistics
    - % nfsstat –s (display statistics of NFS server)
    - % nfsstat –c (display statistics of NFS client)

```
cshome:~ -lwhsu- nfsstat -s

Server Info:
  Getattr   Setattr    Lookup  Readlink      Read     Write    Create    Remove
122903379  44946452 236493933     86636 207570824 123929056   4639946   4706741
   Rename      Link   Symlink     Mkdir     Rmdir   Readdir  RdirPlus    Access
  1452831    851435     14512    474270    400621   1551728    763208 349261427
    Mknod    Fsstat    Fsinfo  PathConf    Commit
    40152  27594331     33528       257  87440831
Server Ret-Failed
       125315642
Server Faults
             0
Server Cache Stats:
   Inprog      Idem  Non-idem    Misses
        0       319         0   4343904
Server Write Gathering:
 WriteOps  WriteRPC   Opsaved
123928555 123929056       501
```

Components of NFS – NFS Utilities (2)

- showmount
  - % showmount –e cchome
    - show the hosts's export list
  - % showmount –a
    - List all mount points

```
bsd4:~ -lwhsu- showmount -e csdisk0
Exports list on csdisk0:
/disk0/distfiles                    140.113.235.0/255.255.255.0
/disk0/alpha                        140.113.235.0/255.255.255.0
/disk0/ISO                          140.113.235.0/255.255.255.0
```

```
showcsdisk0 [~] -lwhsu- showmount -a
All mount points on csdisk0:
140.113.168.126:/disk0/alpha
140.113.17.0/255.255.255.0:/disk0/linuxhome
140.113.17.41:/disk0/linuxhome
140.113.17.41:/disk0/linuxhome/packages
140.113.209.0/255.255.255.0:/disk0
140.113.209.63:/disk0/WWW
140.113.23.0/255.255.255.0:/disk0/ISO
140.113.23.20:/disk0/ISO
140.113.235.0/255.255.255.0,140.113.235.102:/disk0/alpha
...
```

NFS in FreeBSD

- NFS server
  - Edit /etc/rc.conf

```
…
nfs_server_enable="YES"
nfs_server_flags="-u –t –n 4"
…
```

- NFS client

```
…
nfs_client_enable="YES"
…
```

Automatic mounting

- Problems of /etc/fstab
  - Maintenance of /etc/fstab in large network
  - Crashed NFS server will make operation blocked
  - Crashed NFS server will make other local partitions unavailable
- automount daemon
  - Mount filesystems when they are referenced and unmount them when they are no longer needed
  - Supply a list of replicated filesystems to replace important but crashed NFS servers
  - Transparent to users
- Products
  - automount (from SUN Micro), simple and concise
  - amd (from Jan-Simon Pendry), complicated but more powerful

automount (1)

- Three kinds of configuration files (map)
  - Direct map
  - Indirect map
  - Master map

  Provide information about filesystems that are to be automounted
  - List which direct and indirect maps that automount should pay attention to
  - Difference between direct and indirect
    - All mount points in indirect map has common directory defined in master map

automount (2)

- Example of automount maps
  - Master

```
/net    auto.net        -rw, intr
/-      auto.direct     -ro, intr
```

  - Indirect

```
WWW     -rw,soft,nosuid,vers=2 vega:/home/www
mail    -rw,soft,nosuid,quota  ccserv:/spool/mail
ftp     -ro,soft,nosuid ftp:/home/ftp
```

  - direct

```
/vlsi/vlsi1     -rw,soft,nosuid         scorpio:/vlsi1
/vlsi/vlsi2     -rw,soft,nosuid         scorpio:/vlsi2
```

automount (3)

- Master map
  - /etc/auto.master (Linux)
  - /etc/auto_master (Solaris)
- Restart automounter when you change the maps
  - /etc/init.d/autofs {start | stop} (Solairs)
  - /etc/init.d/autofs {start | stop | reload | status} (Linux)

automount (4)

- Replicated filesystem
  - There are several identical NFS and I would like to mount anyone of them
  - Constrain
    - Read-only
    - These replicated filesystem should be truly identical
  - Automounter will choose a server based on its own idea of which one is the best

```
/usr/man        -ro     chimchim:/usr/man band:/usr/man
/www/data       -ro     ccbsd4,altair:/www/data
```

amd (1)

- Advantages over automount
  - Sends “keep alive” queries to remote servers at regular intervals and maintains a list of servers that are accessible
  - Return an “operation would block” rather than hanging
  - Not proprietary source code
  - Offer another mount types that are not supported by automount
  - Map syntax is more generic
  - Provide a query-and-manipulation tool, amq
  - …

amd (2)

- Flexible map syntax
  - One map used by many machines
  - Contain conditions that control which parts of map entry are activate
    - Selector variable

```
/defaults       type:=nfs;fs:=${autodir}/${key};opts:=nfsv3,rw, \
                grpid,quota,intr,soft,nodev,nosuid,resvport, \
                timeo=10,retrans=5,nqnfs
mail            rhost:=ccserv;rfs:=/spool/mail
ftp             rhost:=ftp;rfs:=/home/ftp
raid1           host==cchome;type:=ufs;dev:=/dev/da0s1e\
                host!=cchome;type:=nfs;rhost:=cchome;rfs:=/${key};\
                opts:=nfsv3,rw,grpid,soft,nodev,nosuid,resvport
drongo          host==magpie;type:=link;fs:=/${key} \
                host!=magpie;type:=nfs;rhost:=magpie;rfs:=/${key}
```

amd (3)

Selector	Description
arch	Architecture of the current machine
autodir	Default directory under which to mount filesystems
domain	Local NIS domain name
host	Local hostname
key	Volume name being resolved
map	Name of mount map being used
os	Operating System

Option	Description
rhost	Remote host on which the volume lives
rfs	Remote filesystem name
type	Type of mount, nfs or ufs (local disk)
fs	Local mount point
opts	Mount options
remopts	Options to use if server is nonlocal

amd (4)

- Starting amd
  - # amd -a /tmp_mnt -l syslog -x fatal, error, user /net auto.home
  - /etc/rc.conf:
    - amd_enable="YES"
  - /etc/rc.d/amd start
- Stopping amd
  - # kill -15 <amd_pid>
  - /etc/rc.d/amd stop

options	Description
-x	Sets run-time logging options, such as fatal, error, user, warn, info, …
-r	Restart existing mounts
-l	Log file name or “syslog”
-a	Specify alternative location for mount points
/net	Sets the automount directory
auto.home	The map files

amd (5)

- Remount without kill amd
  - Unmount such mounted partition
    - # umount /amd/magpie
  - Delete such virtual /net/DIR
    - # rm /net/magpie
  - cd /net/DIR
    - # cd /net/magpie
- amd in FreeBSD
  - /etc/rc.conf

```
amd_enable="YES"
amd_flags="-a /amd -c 1800 -d cs.nctu.edu.tw -l /var/log/amd.log -x all /net auto.home"
```