

Mirror Mounts

Tom Haynes

thomas.haynes@sun.com

<http://blogs.sun.com/tdh>

Problem Statement

- Want client to automatically mount filesystems as we detect a new filesystem

Automatically as in ...

- Only explicit use of the mount(1M) command is to get the starting point
 - > But user could mount inside the space
- Only explicit use of the umount(1M) command is the starting point
 - > But user can umount(1M) mirror mounts if they desire

Why mirror mounts?

- Proliferation of server exports
- Automounter issues
- Ease of administration

Proliferation of exports

- With zfs, new datasets are cheap to create
 - > Ordinary users may create with RBAC roles
 - > Easy to deeply nest
 - > Can be automated
- Admin may be unaware of new filesystems

Automounter issues

- Need permissions to edit maps
- Even in */net*:
 - > Changes are slow to propagate
- With mirror mounts, the next READDIR can detect the new share

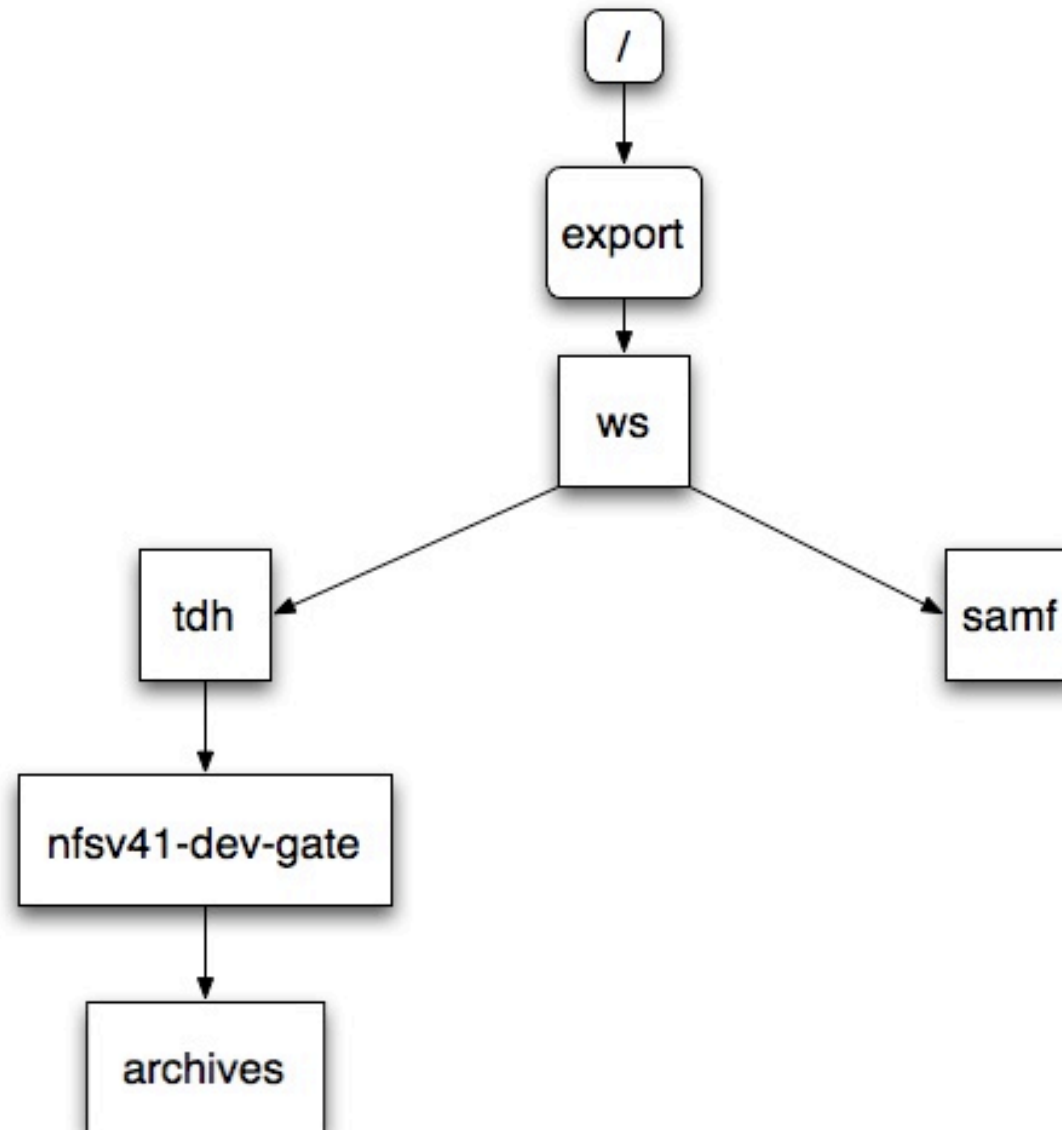
Ease of administration

- Mirror mount inherits properties of parent
 - > Any command line overrides
 - > Any security options
- Follows automount model
 - > umount can unload them
 - > Mirror mounts expire
 - > Default is 600s
 - > Can be modified by parent mount
- Has some new features
 - > umount is recursive

Setting up a server's namespace

```
# zpool create -f tank /dev/dsk/c1d1s3
# zfs create tank/ws
# zfs set sharenfs=rw,anon=0 tank/ws
# zfs set mountpoint=/export/ws tank/ws
# zfs create tank/ws/tdh
  ... share options are inherited
# zfs create tank/ws/samf
# zfs create tank/ws/tdh/nfsv41-dev-gate
# zfs create tank/ws/tdh/nfsv41-dev-gate/archives
```


Resulting Filesystem Hierarchy



Start of client interaction

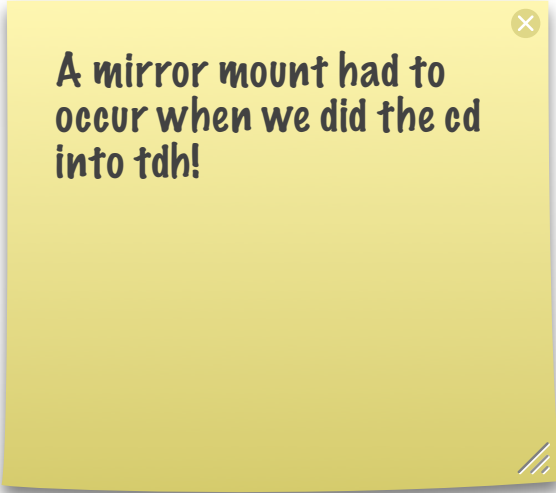
```
# showmount -e sunnfsv4-9
export list for sunnfsv4-9:
/export/ws/samf                (everyone)
/export/ws/tdh                 (everyone)
/export/ws/tdh/nfsv41-dev-gate/archives (everyone)
/export/ws/tdh/nfsv41-dev-gate (everyone)
/export/ws                     (everyone)
# mkdir /mms
# mount sunnfsv4-9:/ /mms
# cd /mms/export/ws
# ls -la
total 11
drwxr-xr-x  4 root    root          4 May 12 08:21 .
drwxr-xr-x  4 root    sys          512 May 12 08:21 ..
drwxr-xr-x  2 root    root          2 May 12 08:21 samf
drwxr-xr-x  3 root    root          3 May 12 08:22 tdh
```

Prior to mirror mounts

```
# uname -a
SunOS sunnfsv4-3 5.10 Generic_120012-14 i86pc i386 i86pc
# mount sunnfsv4-9:/ /mms
nfs mount: sunnfsv4-9://: Permission denied
# mount sunnfsv4-9:/export/ws /mms
# cd /mms
# ls -la
total 11
drwxr-xr-x   4 root      root           4 May 12 08:21 .
drwxr-xr-x  40 root      root        1024 May 12 08:39 ..
drwxr-xr-x   2 root      root           2 May 12 08:21 samf
drwxr-xr-x   2 root      root           2 May 12 08:21 tdh
# cd tdh
# ls -la
total 6
drwxr-xr-x   2 root      root           2 May 12 08:21 .
drwxr-xr-x   4 root      root           4 May 12 08:21 ..
```

With mirror mounts

```
# uname -a
SunOS sunnfsv4-4 5.11 snv_85 i86pc i386 i86pc
# cd tdh
# ls -la
total 9
drwxr-xr-x  3 root      root          3 May 12 08:22 .
drwxr-xr-x  4 root      root          4 May 12 08:21 ..
drwxr-xr-x  3 root      root          3 May 12 08:22 nfsv41-dev-gate
```

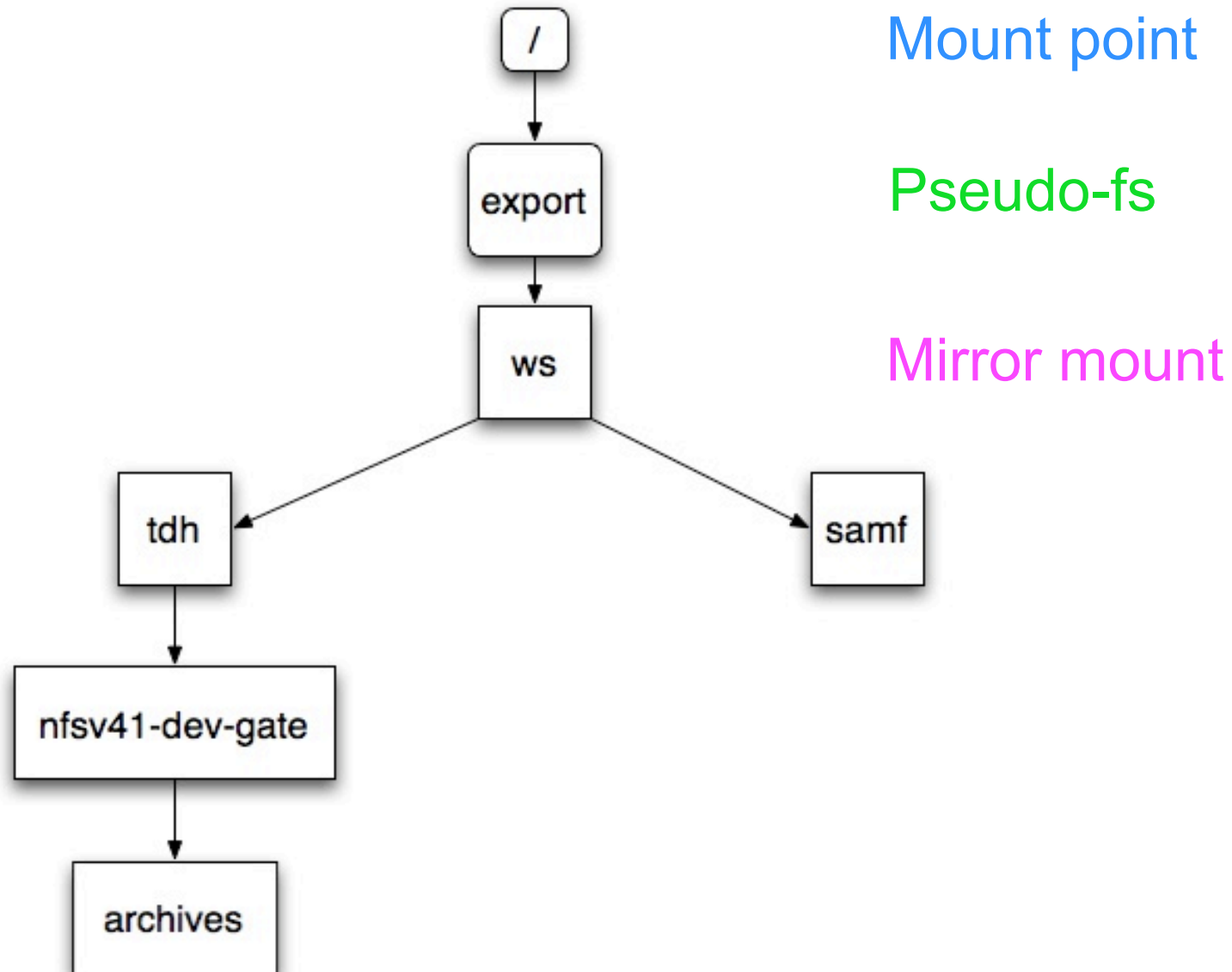


A mirror mount had to occur when we did the cd into tdh!

Is it a mirror mount?

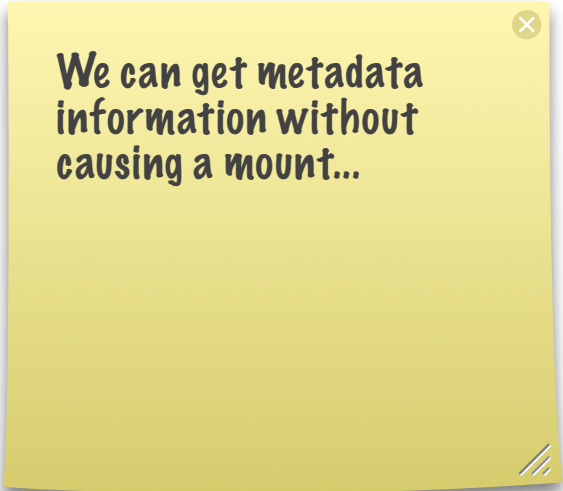
```
# nfsstat -m `pwd`  
/mms/export/ws/tdh from sunnfsv4-9:/export/ws/tdh  
Flags:  
vers=4,proto=tcp,sec=sys,hard,intr,link,symlink,acl,mirrormount,  
rsize=1048576,wsiz=1048576,retrans=5,timeo=600  
Attr cache:      acregmin=3,acregmax=60,acdirmin=30,acdirmax=60  
  
# cd ..  
# nfsstat -m `pwd`  
/mms/export/ws from sunnfsv4-9:/export/ws  
Flags:  
vers=4,proto=tcp,sec=sys,hard,intr,link,symlink,acl,mirrormount,  
rsize=1048576,wsiz=1048576,retrans=5,timeo=600  
Attr cache:      acregmin=3,acregmax=60,acdirmin=30,acdirmax=60  
  
# cd ..  
# nfsstat -m `pwd`
```

Why is ws a mirror mount?



When does a mirror mount occur?

```
# pwd
/mms/export/ws/tdh
# ls -la nfsv41-dev-gate
total 9
drwxr-xr-x  3 root      root          3 May 12 08:22 .
drwxr-xr-x  3 root      root          3 May 12 08:22 ..
drwxr-xr-x  2 root      root          2 May 12 08:22 archives
# nfsstat -m | grep archives
# nfsstat -m | grep gate
/mms/export/ws/tdh/nfsv41-dev-gate from sunnfsv4-9:/export/ws/tdh/
nfsv41-dev-gate
```



**We can get metadata
information without
causing a mount...**

Why don't we load on a readdir?

- It costs ...
- Consider a home directory with 1000s of accounts
- Consider share access lists with 100s of hosts
- Look at:
 - > Cthon 06: Scaling NFS Services
 - > Cthon 07: The Management of Shares

When does the umount occur?

- If idle, after a default of 600s
- When the user manually unloads it

/etc/default/autofs

```
# The duration in which a file system will remain idle before being  
# unmounted. This is equivalent to the "-t" argument to automount.  
#AUTOMOUNT_TIMEOUT=600
```

Yank 'em out

```
# mount sunnfsv4-9:/ /mms
# cd /mms/export/ws/samf/
# cd ../tdh/nfsv41-dev-gate/archives/
# cd
# nfsstat -m | grep mirror | wc -l
    5
# umount /mms
# nfsstat -m | grep mirror | wc -l
    0
```

How does it work?

- At the vnode layer
- We detect a change in filesystem id
- We load a new vnodeops array for the NFSv4 operations
- On a per VOP basis:
 - > Decide if we need to cross filesystem boundary
 - > If so, mount the new filesystem
 - > OTW with NFSv4 operations
 - > No use of the MOUNT protocol
- See [uts/common/fs/nfs/nfs4_stub_vnops.c](#)

Call Outs

- Calum Mackay
- Helen Chao
- Lily Li
- Bill Baker
- Robert Thurlow
- Rich Brown
- Evan Layton
- Alok Aggarwal