NFS Observability: The Undiscovered Country

Bill Baker

Senior Staff Engineer

Sun Microsystems

Bill.Baker@Sun.com

NFS Observability

The Undiscovered Country?

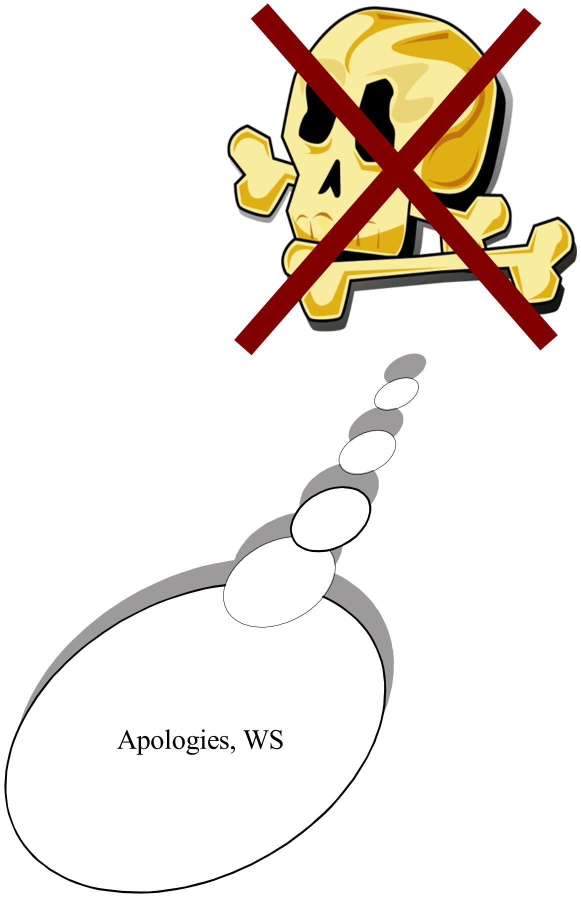
"The undiscovered country from whose bourn no traveller returns, puzzles the will and makes us bear those ills we have than fly to others that we know not of?"

Hamlet – Act III, Scene 1

NFS Observability

The Undiscovered Country

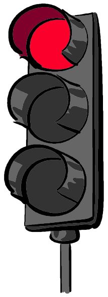
- Ghost of Observability Past
- Ghost of Observability Present
- Roll your own, today!
- The Future



NFS Observability

Ghost of Observability Past

- “NFS Server not responding” ?!?
- nfsstat
  - -c, -s, -m, -r

- Snoop
  -  Protocol experts only!

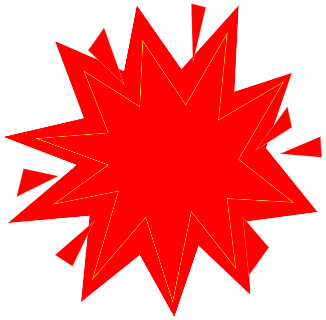
NFS Observability

Ghost of Observability Present

- Server side v4 state per client
  - /var/nfs/v4_state
- Internal “fact sheets”
  - echo ::nfs4_diag | mdb -k
- Delegation statistics
  - kstat -n nfs4_callback_stats

NFS Observability

Roll your own, today!



Dtrace

NFS Observability

Roll your own, today!

- Why are my files owned by nobody?

```
sdt::nfs_idmap_str_uid:nfs4-str-uid {
      printf(“%s”, stringof(arg0));
}
```

- Why am I getting “permission denied?”

```
fbt::geterrno4:entry / arg0 == 10039 / {
      printf(“ID mapping domain mismatch\n”);
}
```

- What's my ID mapping domain?

```
cat /var/run/nfs4_domain
```

NFS Observability

Roll your own, today!

- What are the DNS TXT records?

```
# cat sample.d

#!/usr/sbin/dtrace -s
nfsmapid$1:::thread-domain
{
      printf(“%s”, copyinstr(arg0));
}

# sample.d `pgrep nfsmapid`
```

- Sparc: 120046-01, x86: 120047-01

NFS Observability

Roll your own, today!

- Failed nfs client side opens

```
#!/usr/sbin/dtrace -s

fbt::nfs4_open:entry
{
        self->vp = *(vnode_t **)arg0;
}

fbt::nfs4_open:return
/ self->vp != NULL && arg1 !=  0 /
{
        printf("open %s failed %d\n",
        (string) self->vp->v_path, arg1);
}
```

NFS Observability

Roll your own, today!

- Server write latency

```
#!/usr/sbin/dtrace -Fs
fbt::rfs3_write:entry
{
      self->wresp = (WRITE3res *)arg1;
      self->t0 = timestamp;
}
fbt::rfs3_write:return
/ self->wresp->status == NFS3_OK /
{
      printf("%d bytes written in %d usecs\n",
      self->wresp->res_u.ok.count,
      (timestamp-self->t0)/1000);
}
```

NFS Observability

Roll your own, today!

- Client network latency

```
#!/usr/sbin/dtrace -qs
fbt::rfs4call:entry
{
        self->tag = (string)
        `nfs4_ctags[((COMPOUND4args_clnt *)arg1)->ctag].ct_str;
        self->t0 = timestamp;
}
fbt::rfs4call:return
{
        printf("operation %s done in %d usecs\n", self->tag,
        (timestamp-self->t0)/1000);
}
```

NFS Observability

The Future: General Filesystem Observability

- Fsstat
  - Similar to vmstat
  - Driven by statistics at the vnode/vfs layer
  - fsstat [-t type | mnt-pt] [internal [count]]
  - Statistics aggregated for entire fs
  - I/O throughput, latency, ops-per-second
  - Not a vnode snooper (should it be?)
  - Not per file (should it be?)

Coming soon

NFS Observability

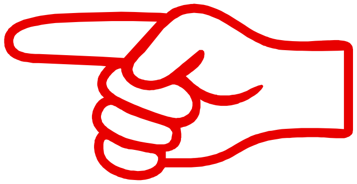
The Future: Better Tools

- nfsrpt - server activity tracing

```
172.20.27.151(V4)  121108  getattr       1       0     0      0   893
172.20.27.151(V4)  121108  open          1       0     0  10011   423
172.20.27.151(V4)       0  setclientid   1       0     0      0   469
172.20.27.151(V4)       0  sclntid_conf  1       0     0      0   2456
172.20.27.151(V4)  121108  open          1       0     0      0   988
172.20.27.151(V4)  121108  open_confirm  1       0     0      0   362
172.20.27.151(V4)  121108  read          1       3     0      0   783
172.20.27.151(V2)  121108  root          1       0     0      0   304
172.20.27.151(V2)  121108  readdir       1       0     0      0   752
172.20.27.151(V2)  121108  lookup        1       0     0      0   251
```

- nfstop - hot clients & hot files

- /usr/demo/dtrace



Watch this space

NFS Observability

The Future: An NFS provider

- What & Why?
  - Abstracts scripts from the probes
  - Potentially creates friendlier presentation
- Suggestions?
  - nfs-discuss@opensolaris.com

NFS Observability

The Undiscovered Country

- Historically very weak
- Some improvements in v4
  - Still pretty weak
- Roll your own with Dtrace
  - Handy, but …
- More, better tools
  - Up-level the data for administrators