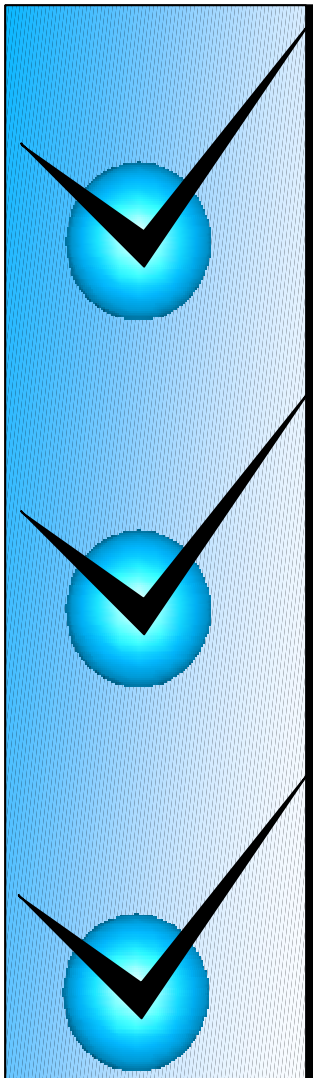


# NFS: A Service Provider's Perspective

Tim Bosserman

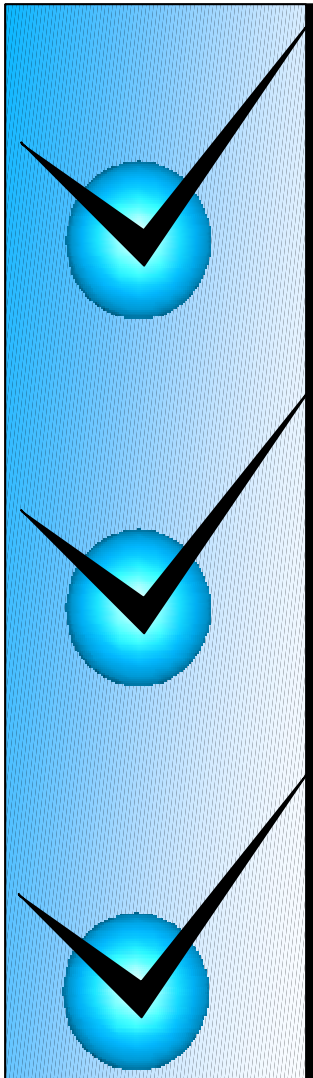
Consulting Research Engineer

Earthlink, Inc.

tboss@corp.earthlink.net

# Discussion Topics

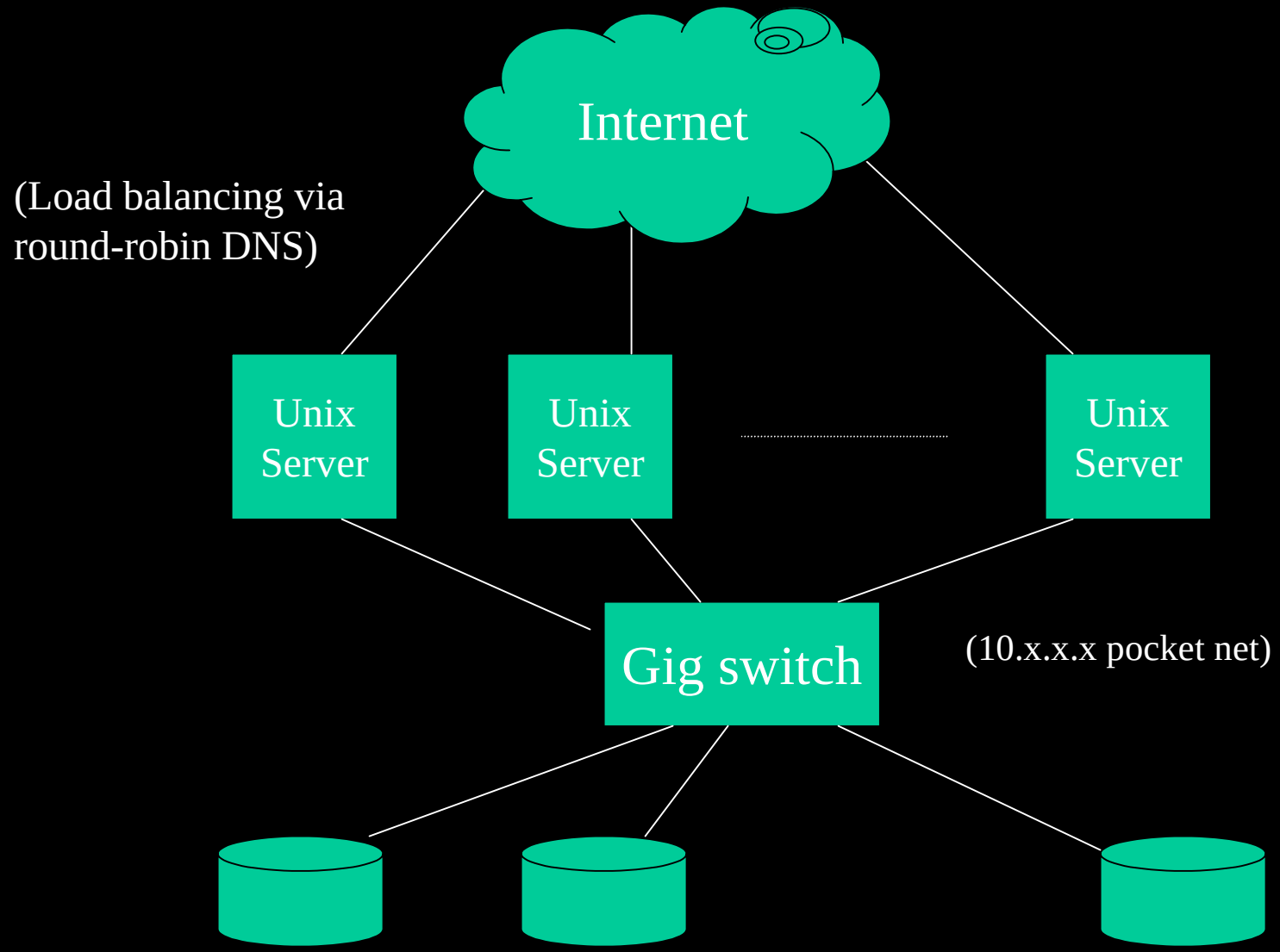
- Why NFS?
  - The “EarthLink Standard Architecture”
- NFS Drawbacks
- File Locking
- Future Technologies
  - DAFS: The Direct Access FileSystem
  - NFS Version 4



# Why NFS?

- Stateless Architecture
- Redundancy
- Scalability
- Vendor Agnostic
- Widely Supported
- Cost Effective

# EarthLink Standard Architecture



# NFS Drawbacks

- Inconsistent Implementations
- Performance Bottlenecks
- Attribute caching
- Mmap()
  - GNU grep on a rapidly changing file
- File Locking

# Performance Bottlenecks

- Exhausting kernel resources
  - cltoomany symptom
  - Overflowing kernel queues
- Latency
- Large file read/write contention

# Workarounds / Tuning

- Flushing attribute cache
  - Utime() or utimes() with NULL time
- Balancing mount points
  - TCP vs. UDP
  - NFS v2 vs. v3
- Squeezing pathnames (getattr ops)
  - /data/vol/vol1/spam/username vs.
  - /data/vol1/spam.username
- Reduce size of name cache

# File Locking

- Single biggest problem with NFS
- Dubbed: Network Failure System
- Not implemented by all operating systems (Mac OS X, Tru64 version 4)
- Badly implemented by others
- Multiple processes may be granted same exclusive lock
- No processes will be granted lock to unlocked file

# Code Fragment

```
for (i = 0; i < num_children - 1; ++i)
    if (fork() == 0)
        break;

for (;;)
{
    filenum = random() % numfiles;
    if (lockf(fds[filenum], F_LOCK, 0) < 0)
        errout();
    if (lockf(fds[filenum], F_ULOCK, 0) < 0)
        errout();
}
```

# Workarounds

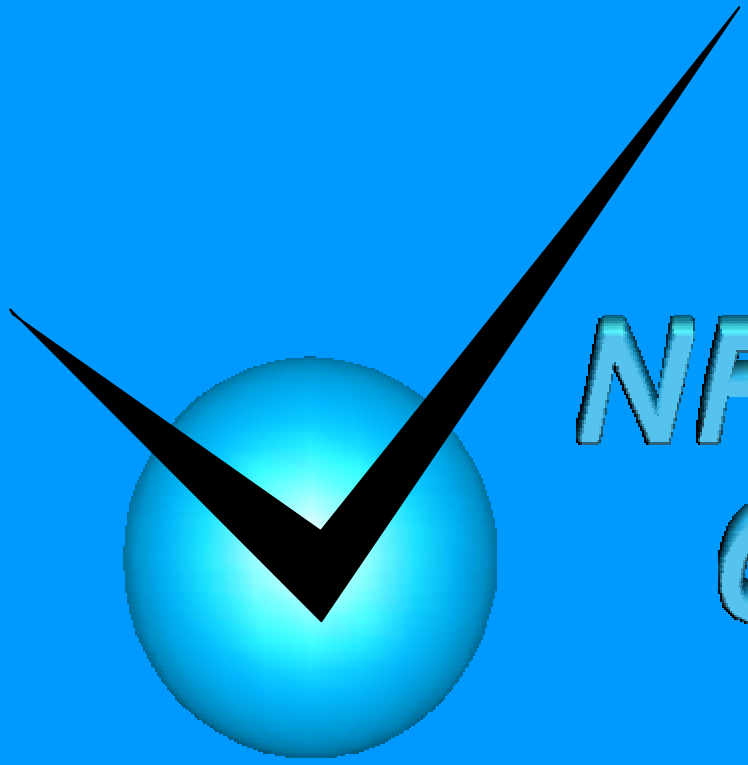
- Rewrite code to not use locking
  - Example: maildir mailbox format
- Use the filesystem
  - open(fname, O_EXCL|O_CREAT)
  - link()
- Write your own lock manager
- Parallel directory structure on local disk

# Upcoming Technology

- NFS v4
- DAFS: The Direct Access Filesystem

# DAFS

- Highspeed “cousin” to NFS
  - Borrows heavily from NFS in places
- Focused on “narrow sharing” environment
- Two ways to implement
  - User-level library
    - Highest possible performance
    - Programs must be modified
  - Kernel virtual filesystem
    - Transparent to user programs



# NFS Vendors Conference