# ngnfs
## block cache coherence

Zach Brown

<zab@zabbo.net>, <zab@versity.com>

https://git.infradead.org/?p=users/zab/ngnfs-progs.git

# Use case: multi-node POSIX FS as staging buffer

Archive:

- Files stored in FS via services (NFS/samba/S3)
- Archiving agent reads from FS and writes to archive
- File data truncated ("offline" flag)

Recall:

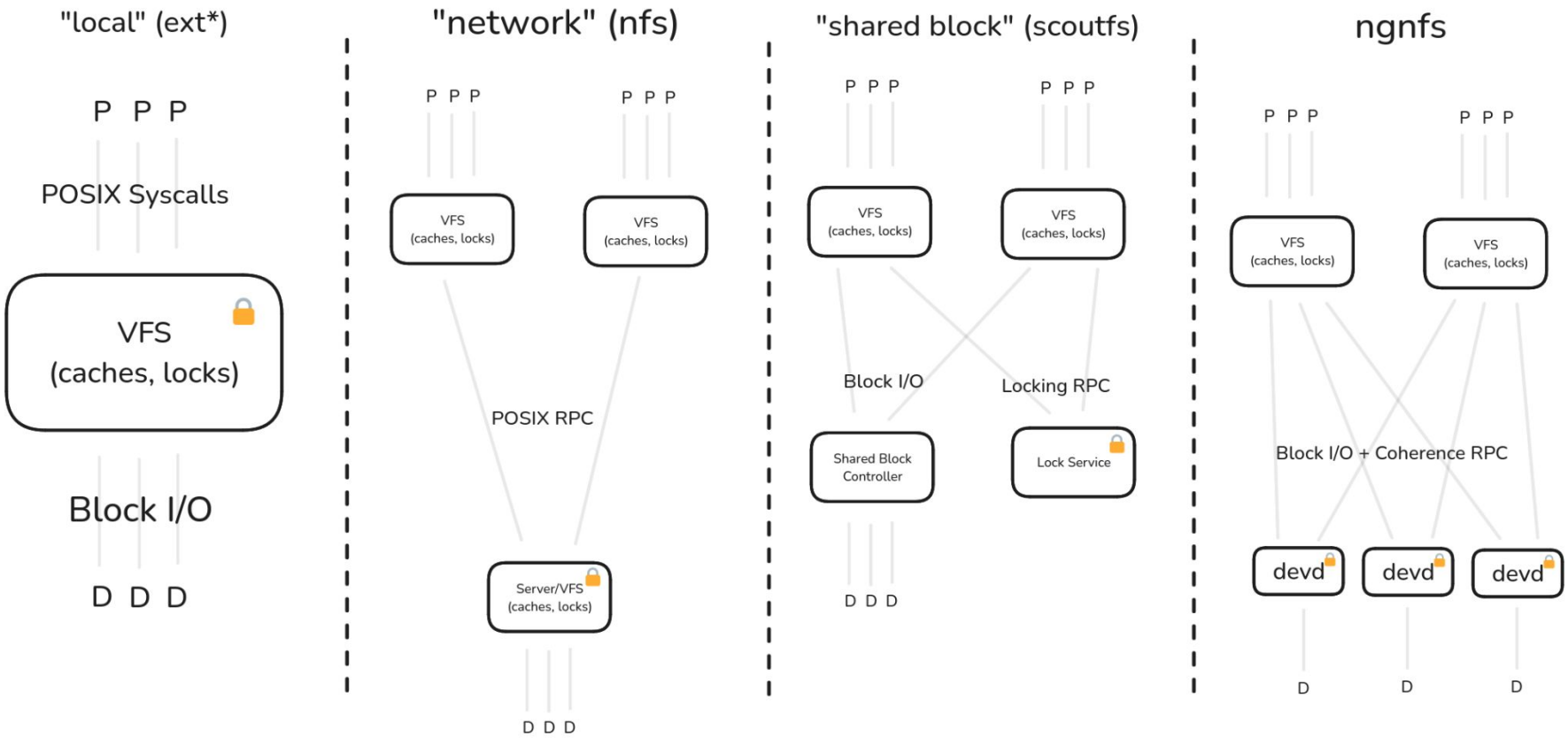
- Services block reading from FS
- Agent reads from archive and writes to FS
- Service reads complete

Environment:

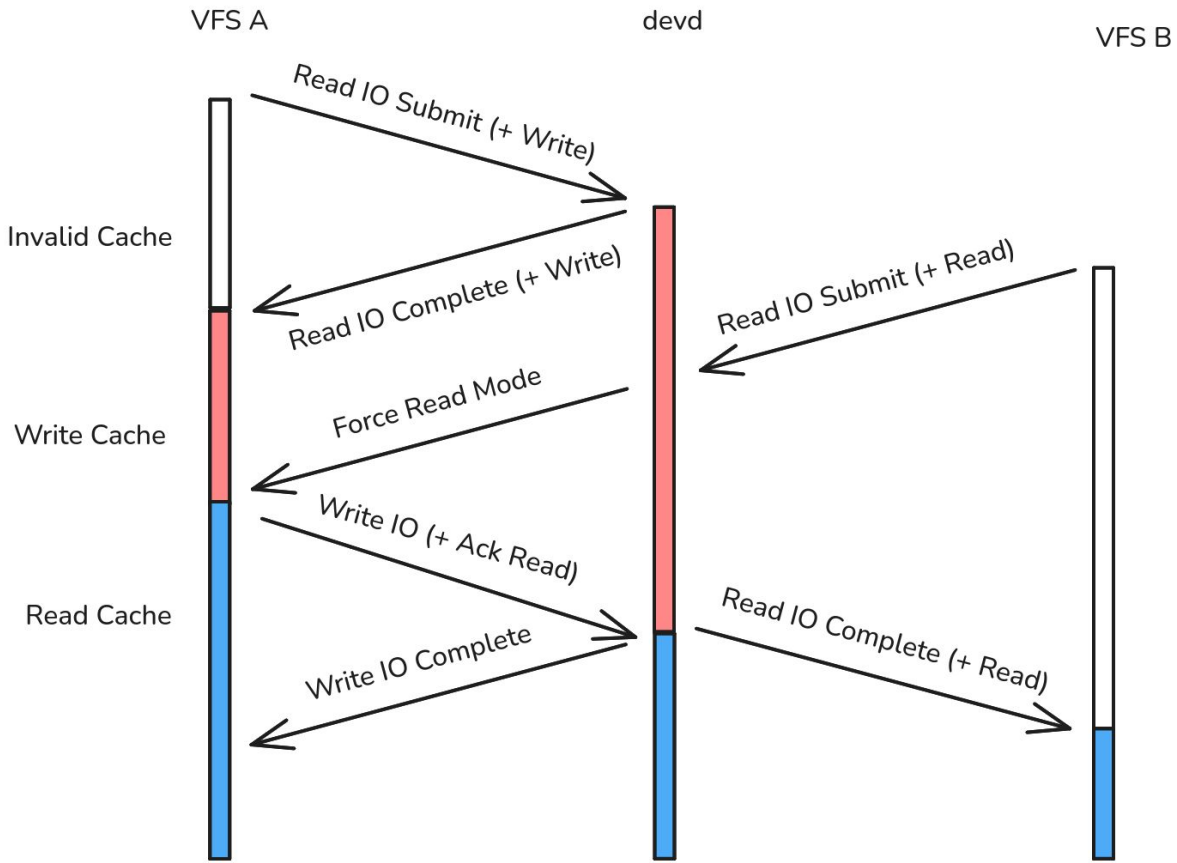
- Billions of files
- Many nodes for aggregate bandwidth
- Non-uniform archive access
- Example Node: 2 100Gb eth, 8 32Gb FC, 2 100Gb EDR IB


*Tape's demise has been greatly exaggerated.*

# Sampling of consistency models



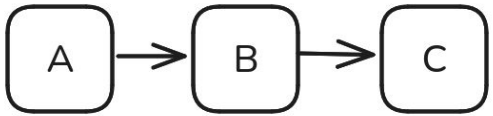
# ngnfs block IO + cache coherence



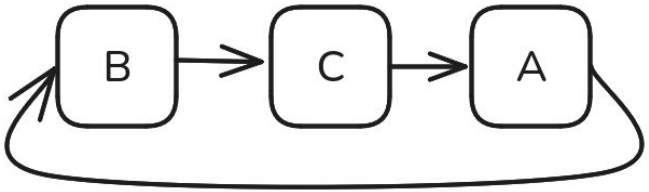
# Cross-directory rename can be globally serialized.

*/* "Perverted" doesn't begin to describe it. */*

$ mkdir -p a/b/c



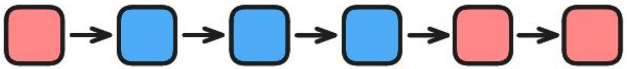

$ mv a a/b/c/a
mv: cannot move 'a' to a subdirectory of itself, 'a/b/c/a'



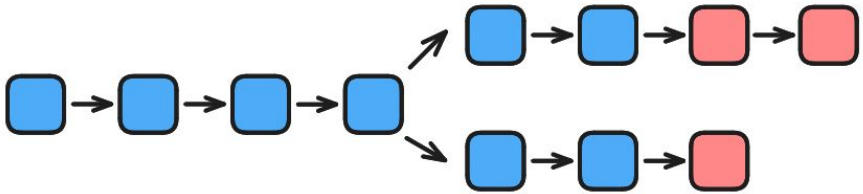

```
struct dentry *lock_rename(struct dentry *p1, struct dentry *p2)
{
    /* ... */
    mutex_lock(&p1->d_sb->s_vfs_rename_mutex);
    return lock_two_directories(p1, p2);
}
```

# ngnfs block transactions naturally pin parent walk.


Return error with read-only access



Successfully rename to remote ancestor



Successfully rename to distant cousin

# Q/A

Zach Brown

<zab@zabbo.net>, <zab@versity.com>

https://git.infradead.org/?p=users/zab/ngnfs-progs.git