



## Restart/Continue pNFS Metastripe '90



## Metastripe: pNFS for Metadata

- Adds cooperating MDS servers, scale-out metadata, and parallel metadata operations to NFSv4, using new pNFS metadata layout type(s)
- Adapts and extends the proposal by Eisler, in 2010
- First presented at NFSv4 WG IETF 86, Orlando



## Recap: changes from prior drafts

- Reduced layout state overhead
  - Avoid requirement to take metastripe layouts on regular files
  - Propose *stripe hints* (attribute)
    - Coarse-grained layout (file-system)
    - Deviceid hint (attribute)



## Filehandle Striping (new name!)

- Filesystem “singleton” layout
  - Holds device list and stripe pattern
  - Permits clients to request stripe hint
    - Recommended attribute
  - Used for:
    - Open, GETATTR, LOCK



## Directory Striping (new name!)

- All other metastripe (layouts on directories)
  - Essentially just like original metastripe
  - Used for:
    - Name-based operations (CREATE)
    - Directory enumeration (in parallel)



## Simplify metadata layout slightly

- Remove layout filehandles (try to)
- Simplified device model
  - Define one device structure and layout device presentation
    - always use it
- Keep “offsets” opaque



## Changes since 01

- Layout sub-type names changed
  - Structure and semantics unchanged
- Improved language in several areas
- New directory striping algorithm (CEPHFRAG) added



## Recent Changes

- Opaque data in LAYOUTGET (still needed)
- New directory striping algorithm (CEPHFRAG)





## MDN\_ALG\_CEPHFRAG

- The CEPHFRAG algorithm describes the Ceph algorithm for placing new directory entries on “fragments”
  - A striping algorithm based on recursive hashing and splitting
  - Shows generality of the mechanism, frag trees are typed seed data already provided for
- These changes are now published.



- Ganesha
  - Provisionally complete, WIP source available
    - <https://github.com/linuxbox2/nfs-ganesha> (metastripe)
- PyNFS
  - Nearly complete set of initial tests, WIP source available
    - <https://github.com/linuxbox2/pynfs> (master)



## Current Draft

<http://tools.ietf.org/html/draft-mbenjamin-nfsv4-pnfs-meta-stripe-03>

## Source

<https://github.com/linuxbox2/metastripe>



Q/A