draft-ietf-nfsv4-pnfs-obj-01 updates

Benny Halevy

bhalevy@panasas.com

July 11, 2006

66’th IETF - Montreal

Preface

- This presentation summarizes the major additions and changes in draft-ietf-nfsv4-pnfs-obj-01.txt

What's changed in -01

- Editing-level changes
- `pnfs_osd_addr_type4`
- `pnfs_osd_layout_update4`
- `pnfs_osd_ioerr4`
- `pnfs_layouthint4`
- Object-Based Layout Segments
- LAYOUTCOMMIT
- Recalling Layouts
- Security Considerations

Editing-level changes

- Text reorganized to help readability
- Explanatory text added for internal data types to help understand and implement the draft
- **Naming convention:** types begin with `pnfs_osd`
- References to external data types scrubbed and corrected where needed

pnfs_osd_addr_type4

```
struct pnfs_osd_deviceaddr4 {
   union target switch (pnfs_osd_addr_type4 type) {
      case OBJ_TARGET_NETADDR:
          pnfs_netaddr4   netaddr;
      case OBJ_TARGET_IQN:
          string          iqn<>;
      case OBJ_TARGET_WWN:
          string          wwn<>;
      default:
          void;
   };
   uint64_t            lun;
   opaque              root_id<>;
};
```

- Allows specifying a `pnfs_netaddr4` instead of a `string` containing an ip address.
- Structure union-ified for simplicity.

pnfs_osd_layout_update4

```
struct pnfs_osd_layoutupdate4 {
    pnfs_osd_deltaspaceused4    delta_space_used;
    pnfs_osd_ioerr4             ioerr<>;
};
```

- Removed `time_metadata` as it's redundant with `LAYOUTCOMMIT4args.time_modify`

pnfs_osd_ioerr4

```
struct pnfs_osd_ioerr4 {
    pnfs_osd_objid4     component;
    length4             offset;
    length4             length;
    bool                iswrite;
    pnfs_osd_errno4     errno;
};
```

- Added "`bool iswrite`" field to `pnfs_osd_ioerr4` to signify errors on data modifying osd operations.

pnfs_layouthint4

```
union num_comps_hint4 switch (bool valid) {
     case TRUE:
         uint32_t            num_comps;
     case FALSE:
         void;
};
. . .
struct pnfs_osd_layouthint4 {
    num_comps_hint4         num_comps_hint;
    stripe_unit_hint4       stripe_unit_hint;
    group_width_hint4       group_width_hint;
    group_depth_hint4       group_depth_hint;
    mirror_cnt_hint4        mirror_cnt_hint;
    raid_algorithm_hint4    raid_algorithm_hint;
};
```

- Type beefed up with all map parameters, all are optional.
- Added usage text.

Object-Based Layout Segments

- Object based layout segments cannot be merged or split due to immutability of OSD capabilities.
- New text specifies rules for handling byte ranges in `CB_LAYOUTRECALL` and `LAYOUTEGET`
- Key difference from file/block layouts is that client needs to keep track of each segment it got.
  - Client should return all segments overlapping with the recalled range (and matching iomode etc.)
  - Client MUST NOT return a sub-range of a layout segment.

LAYOUTCOMMIT

- Added explanatory text about non-usage of `LAYOUTCOMMIT`'s byte range. Reserved for future use, must be set to 0 for now.
- Explained usage of byte ranges in `pnfs_osd_ioerr4` which may be sent in `pnfs_osd_layoutupdate4` as part of `LAYOUTCOMMIT`.

Recalling Layouts

- Added explanatory text to describe in what scenarios the object-based metadata server is expected to recall layouts
  - File's security policy changes
  - File's aggregation map changes
  - Sharing conflicts

Security Considerations

- Merged section 15.2 “Object Layout Security” of draft-ietf-nfsv4-minorversion1-03.txt into this document.
- This section can be safely removed now from the nfsv4.1 draft.