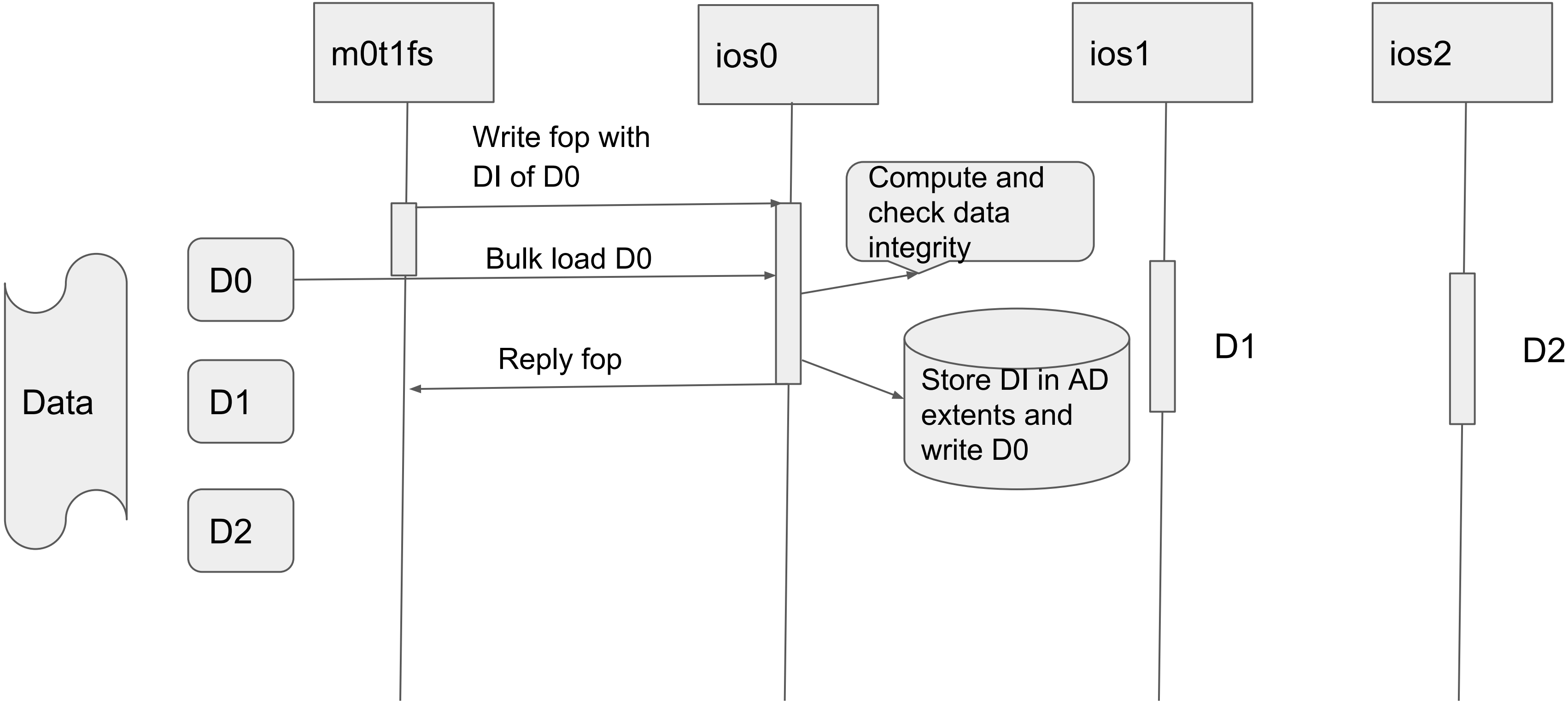
Mero End-to-end Data integrity

By Madhavrao Vemuri

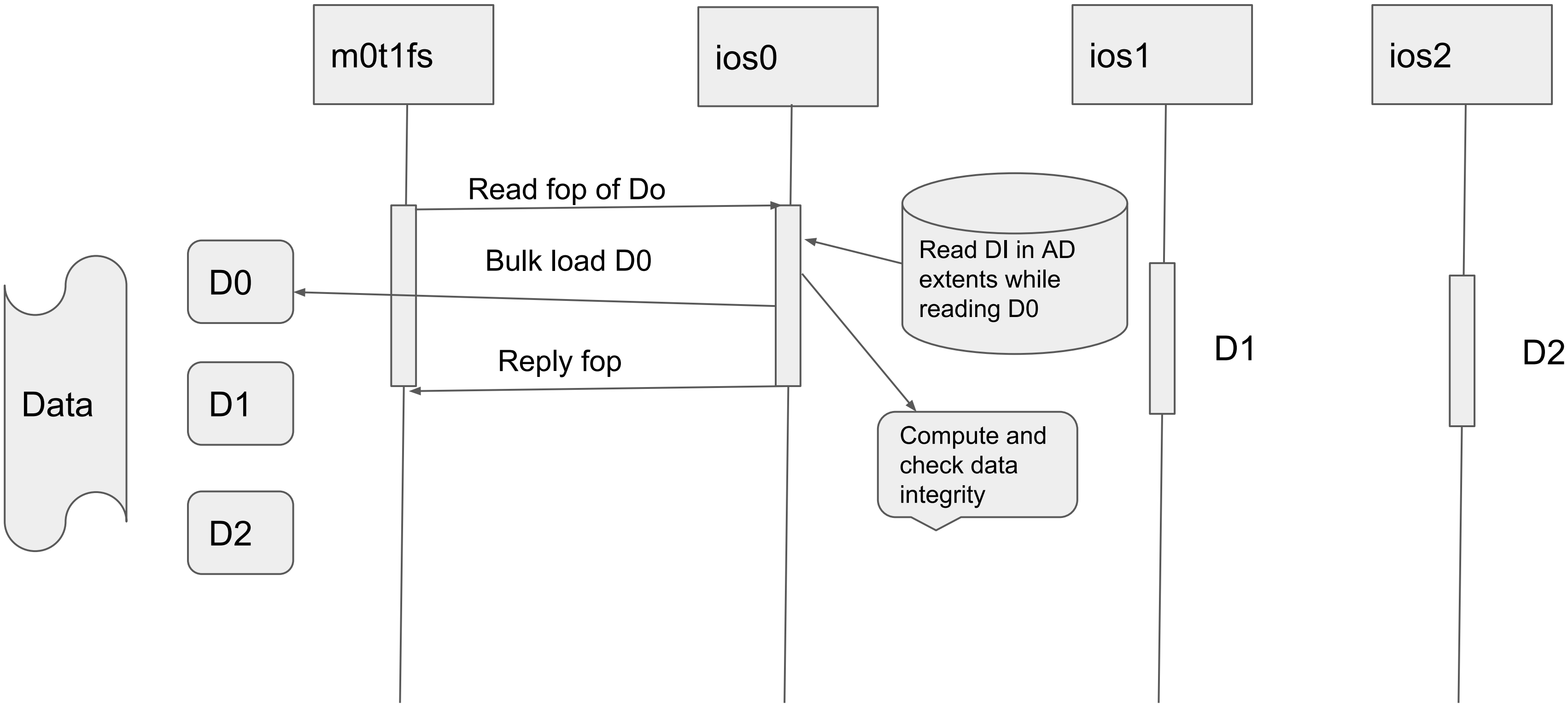
Design Highlights

- Data of each target is divided into blocks of 4096 bytes
- Checksum and tags of 64-bit each for these blocks are computed at m0t1fs (clovis or sns) and sent over wire
- Checksum for data blocks is computed based on checksum algorithm Selected from configuration
- Data integrity type and operations are initialised in m0_file
- Using do_sum(), checksum values are computed for each block of data And using do_chk(), checksum values are verified

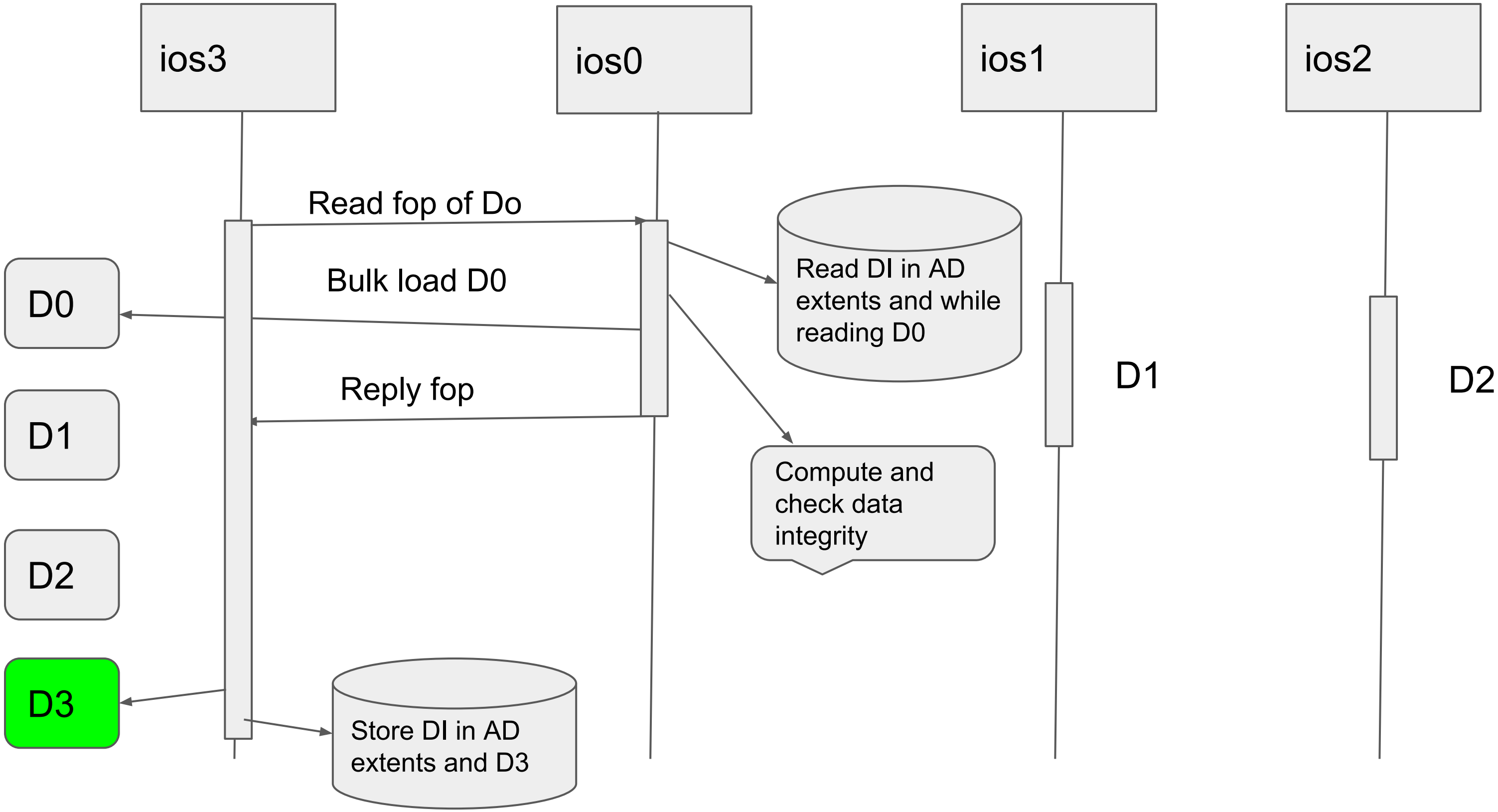
Write


m0t1fs
ios0
ios1
ios2
Write fop with DI of D0
Compute and check data integrity
Bulk load D0
D0
Data
D1
D2
Reply fop
Store DI in AD extents and write D0
D1
D2

Read


m0t1fs
ios0
ios1
ios2
Read fop of Do
Bulk load D0
Reply fop
Read DI in AD extents while reading D0
Compute and check data integrity
Data
D0
D1
D2
D1
D2

SNS



ios3

ios0

ios1

ios2

Read fop of Do

Bulk load D0

Reply fop

D0

D1

D2

D3

Read DI in AD extents and while reading D0

Compute and check data integrity

D1

D2

Store DI in AD extents and D3

Current status

Completed

a) Di is computed at m0t1fs and sent over wire

b) After receiving write fop, checksum is recomputed and verified at ioservice

In progress

a) In be segment block attributes m0_be_emap_seg:ee_battr is added.
And m0_be_emap_seg:ee_val and ee_battr (When b_nob > 0) are stored
In btree.

b) emap split for di data is also work in progress.

c) Write di data to extents while storing the data in disks.
(uses be_emap_split and in place btree insert api’s)

d) Read di data from extents while reading data from disks and verify checksum.

e) In sns while reading data units, verify checksum and while writing, store
di data too.

Checksum Feature Task Break-up

1) HLD (Is it needed ?)
2) DLD ((Is it needed ?))
3) Implementation
    a) Changes in be extmap (Use m0_buf instead of uint64_t) - 2w
    b) Changes in ad, like split and merge - 2w
    c) Check data-integrity in ioservice - 1w
    d) Configuration - Add parameter in conf filesystem for di - 3d
    e) Cluster Testing - 2w
    f) Inspection - 1w