fs - Bug #14800

[ceph-fuse] Fh ref might leak at umounting

02/18/2016 05:22 AM - Zhi Zhang

Status:	Resolved	**Start date:**	02/18/2016
Priority:	Normal	**Due date:**	
Assignee:		**% Done:**	0%
Category:		**Estimated time:**	0.00 hour
Target version:			
Source:	Community (dev)	**Affected Versions:**	
Tags:		**ceph-qa-suite:**	
Backport:		**Component(FS):**	
Regression:	No	**Labels (FS):**	
Severity:	3 - minor	**Pull request ID:**	
Reviewed:			

Description

Recently we meet ceph-fuse hanging issue and have to kill ceph-fuse process to continue. This issue is caused by force umounting ceph-fuse when there are some files being written.

```
...
2016-02-17 12:50:22.188303 7ff6a77fe700  7 client.4338 wrote to 510932646, extending file size
2016-02-17 12:50:22.188312 7ff6a77fe700 10 client.4338 mark_caps_dirty 1000000046f.head(ref=6 ll_r
ef=9 trunc_seq=2 cap_refs={1024=1,4096=1,8192=1} open={2=1} mode=100644 size=510932646/1073741824
mtime=2016-02-17 12:50:22.188312 caps=pAsxLsXsxFsxcrwb(0=pAsxLsXsxFsxcrwb) dirty_caps=Fw objectset
[1000000046f ts 2/0 objects 122 dirty_or_tx 7616166] parents=0x7ff6f0009f70 0x7ff6f0009a10) Fw ->
Fw
2016-02-17 12:50:22.188345 7ff6a77fe700  3 client.4338 ll_write 0x7ff6ac0016b0 510929496~3150 = 31
50
2016-02-17 12:50:22.188839 7ff70eb37780  2 client.4338 unmounting
...
2016-02-17 12:50:22.191395 7ff70eb37780 10 client.4338 check_caps on 1000000046f.head(ref=6 ll_ref
=0 trunc_seq=2 cap_refs={1024=1,4096=0,8192=1} open={2=1} mode=100644 size=510932646/1073741824 mt
ime=2016-02-17 12:50:22.188312 caps=pAsxLsXsxFsxcrwb(0=pAsxLsXsxFsxcrwb) dirty_caps=Fw objectset[1
000000046f ts 2/0 objects 122 dirty_or_tx 7616166] parents=0x7ff6f0009f70 0x7ff6f0009a10) wanted p
AsxXsxFxcwb used Fcb is_delayed=1
2016-02-17 12:50:22.191418 7ff70eb37780 10 client.4338  cap mds.0 issued pAsxLsXsxFsxcrwb implemen
ted pAsxLsXsxFsxcrwb revoking -
2016-02-17 12:50:22.191425 7ff70eb37780 10 client.4338 mark_caps_flushing Fw 1000000046f.head(ref=
6 ll_ref=0 trunc_seq=2 cap_refs={1024=1,4096=0,8192=1} open={2=1} mode=100644 size=510932646/10737
41824 mtime=2016-02-17 12:50:22.188312 caps=pAsxLsXsxFsxcrwb(0=pAsxLsXsxFsxcrwb) dirty_caps=Fw obj
ectset[1000000046f ts 2/0 objects 122 dirty_or_tx 7616166] parents=0x7ff6f0009f70 0x7ff6f0009a10)
2016-02-17 12:50:22.191441 7ff70eb37780 10 client.4338 send_cap 1000000046f.head(ref=6 ll_ref=0 tr
unc_seq=2 cap_refs={1024=1,4096=0,8192=1} open={2=1} mode=100644 size=510932646/1073741824 mtime=2
016-02-17 12:50:22.188312 caps=pAsxLsXsxFsxcrwb(0=pAsxLsXsxFsxcrwb) flushing_caps=Fw objectset[100
0000046f ts 2/0 objects 122 dirty_or_tx 7616166] parents=0x7ff6f0009f70 0x7ff6f0009a10) mds.0 seq
17 used Fcb want pAsxXsxFxcwb flush Fw retain pAsxXsxFxcwb held pAsxLsXsxFsxcrwb revoking - droppi
ng LsFsr
2016-02-17 12:50:22.191467 7ff70eb37780 15 client.4338 auth cap, setting max_size = 0
2016-02-17 12:50:22.191479 7ff70eb37780 10 client.4338 wait_sync_caps want 128 (last is 128, 1 tot
al flushing)
2016-02-17 12:50:22.191497 7ff70eb37780 10 client.4338  waiting on mds.0 tid 128 (want 128)
...
2016-02-17 12:51:08.722751 7ff70eb37780  1 client.4338 dump_cache
2016-02-17 12:51:08.722782 7ff70eb37780  1 client.4338 dump_inode: DISCONNECTED inode 1000000046f
#1000000046f ref 11000000046f.head(ref=1 ll_ref=0 trunc_seq=2 cap_refs={1024=0,4096=0,8192=0} open
={2=1} mode=100644 size=510932646/1073741824 mtime=2016-02-17 12:50:22.188312 caps=pAsxXsx(0=pAsxX
sx) objectset[1000000046f ts 2/0 objects 0 dirty_or_tx 0] 0x7ff6f0009a10)
2016-02-17 12:51:08.722837 7ff70eb37780  2 client.4338 cache still has 0+1 items, waiting (for cap
```

```
s to release?)
2016-02-17 12:51:08.772373 7ff702ffd700 10 client.4338  mds.0 seq now 522
2016-02-17 12:51:08.772387 7ff702ffd700  5 client.4338 handle_cap_grant on in 1000000046f mds.0 se
q 20 caps now pAsxLsXsxFsxcrwb was pAsxXsx
2016-02-17 12:51:08.772420 7ff702ffd700 10 client.4338 update_inode_file_bits 1000000046f.head(ref
=1 ll_ref=0 trunc_seq=2 cap_refs={1024=0,4096=0,8192=0} open={2=1} mode=100644 size=510932646/1073
741824 mtime=2016-02-17 12:50:22.188312 caps=pAsxXsx(0=pAsxXsx) objectset[1000000046f ts 2/0 objec
ts 0 dirty_or_tx 0] 0x7ff6f0009a10) pAsxXsx mtime 2016-02-17 12:50:22.188312
2016-02-17 12:51:08.772469 7ff702ffd700 10 client.4338   grant, new caps are LsFsxcrwb
2016-02-17 12:51:08.772490 7ff702ffd700 10 client.4338 unmounting: trim pass, size was 0+1
2016-02-17 12:51:08.772508 7ff702ffd700 20 client.4338 trim_cache size 0 max 0
2016-02-17 12:51:08.772526 7ff702ffd700 10 client.4338 unmounting: trim pass, size still 0+1
2016-02-17 12:51:13.722943 7ff70eb37780  1 client.4338 dump_cache
2016-02-17 12:51:13.722978 7ff70eb37780  1 client.4338 dump_inode: DISCONNECTED inode 1000000046f
#1000000046f ref 11000000046f.head(ref=1 ll_ref=0 trunc_seq=2 cap_refs={1024=0,4096=0,8192=0} open
={2=1} mode=100644 size=510932646/1073741824 mtime=2016-02-17 12:50:22.188312 caps=pAsxLsXsxFsxcrw
b(0=pAsxLsXsxFsxcrwb) objectset[1000000046f ts 2/0 objects 0 dirty_or_tx 0] 0x7ff6f0009a10)
2016-02-17 12:51:13.723048 7ff70eb37780  2 client.4338 cache still has 0+1 items, waiting (for cap
s to release?)
...
```

The simple steps to reproduce, for example:
1. keep writing a file:
sudo sh -c "tail -f /var/log/ceph/ceph-client.admin.log > /mnt/cephfs/test_file_1"

2. force umount ceph-fuse on another window:
sudo umount -f /mnt/cephfs

3. wait for a while to check ceph-fuse process and log:
ceph-fuse process is still alive but mount point is umounted.

By further digging, we found Fh's ref can't be released in this situation, hence inode's last ref also can't. This is because in ll_open and ll_create, both Fh's and inode's ref will be incremented, then in ll_release, Fh's ref can be reduced to 0 and further reduce inode's ref. But in above situation, ll_release will never be called, so inode's last ref can't be released and inode_map can't be cleared.

The fix that I can think of is like the way in libcephfs, which is to use another set to save the unclosed Fhs (for ceph-fuse use) and clean up them like fd_map (for libcephfs use) at umounting.

History

#1 - 02/18/2016 05:24 AM - Zhi Zhang

https://github.com/ceph/ceph/pull/7686

#2 - 02/22/2016 12:19 PM - Zheng Yan

- Status changed from New to Need Review

#3 - 03/10/2016 06:26 AM - Greg Farnum

- Status changed from Need Review to Resolved