fs - Bug #25113

mds: allows client to create ".." and "." dirents

07/26/2018 01:05 AM - Patrick Donnelly

Status:	Resolved	**Start date:**	07/25/2018
Priority:	Urgent	**Due date:**	
Assignee:	Venky Shankar	**% Done:**	0%
Category:	Correctness/Safety	**Estimated time:**	0.00 hour
Target version:	v14.0.0		
Source:	Development	**Affected Versions:**	
Tags:		**ceph-qa-suite:**	
Backport:	mimic,luminous	**Component(FS):**	MDS
Regression:	No	**Labels (FS):**	task(easy)
Severity:	3 - minor	**Pull request ID:**	
Reviewed:			

Description

Pavani Rajula (our Outreachy Intern) found a fantastic bug. Apparently the MDS will happily allow the client to create "." or ".." directory entries. She found this during work on the CephFS shell. It's not reproducible with ceph-fuse or kclient because Linux catches this at the vfs layer.

CephFS shell command:

```
CephFS:~/>>> mkdir -p a/.
```

(or just call ceph_mkdirs via C.)

From the client log:

```
2018-07-25 20:58:04.185 7f3928b8b700  8 client.4212 _mkdir(#0x1/a, 040775) = 0
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 mkdirs: successfully created directory
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 may_create 0x10000000006.head(faked_ino=0 ref=
3 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.185860 mtime=
2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsxLsXsxFsx(0=pAsxLsXsxFsx) COMP
LETE parents=0x1.head["a"] 0x7f38ec014010); UserPerm(uid: 1163, gid: 1163)
2018-07-25 20:58:04.185 7f3928b8b700 10 client.4212 _getattr mask As issued=1
2018-07-25 20:58:04.185 7f3928b8b700  3 client.4212 may_create 0x7f38ec014010 = 0
2018-07-25 20:58:04.185 7f3928b8b700  8 client.4212 _mkdir(0x10000000006 ., 0775, uid 1163, gid 11
63)
2018-07-25 20:58:04.185 7f3928b8b700 10 client.4212 get_quota_root realm 0x1
2018-07-25 20:58:04.185 7f3928b8b700 10 client.4212 get_quota_root 0x10000000006.head -> 0x1.head
2018-07-25 20:58:04.185 7f3928b8b700 15 inode.get on 0x7f38ec014010 0x10000000006.head now 4
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 get_or_create 0x10000000006.head(faked_ino=0 r
ef=4 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.185860 mti
me=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsxLsXsxFsx(0=pAsxLsXsxFsx) C
OMPLETE parents=0x1.head["a"] 0x7f38ec014010) name .
2018-07-25 20:58:04.185 7f3928b8b700 15 open_dir 0x1afde50 on 0x7f38ec014010
2018-07-25 20:58:04.185 7f3928b8b700 15 inode.get on 0x7f38ec014010 0x10000000006.head now 5
2018-07-25 20:58:04.185 7f3928b8b700 15 client.4212 link dir 0x7f38ec014010 '.' to inode 0 dn 0x1a
f8960 (new dn)
2018-07-25 20:58:04.185 7f3928b8b700 10 client.4212 _mkdir: making request
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 choose_target_mds starting with req->inode 0x1
0000000006.head(faked_ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2
018-07-25 20:58:04.185860 mtime=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=p
AsxLsXsxFsx(0=pAsxLsXsxFsx) COMPLETE parents=0x1.head["a"] 0x7f38ec014010)
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 choose_target_mds inode dir hash is 2 on . =>
```

```
4024326105
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 choose_target_mds 0x10000000006.head(faked_ino
=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.185860
 mtime=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsxLsXsxFsx(0=pAsxLsXsxFs
x) COMPLETE parents=0x1.head["a"] 0x7f38ec014010) is_hash=1 hash=4024326105
2018-07-25 20:58:04.185 7f3928b8b700 10 client.4212 choose_target_mds from caps on inode 0x1000000
0006.head(faked_ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07
-25 20:58:04.185860 mtime=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsxLsX
sxFsx(0=pAsxLsXsxFsx) COMPLETE parents=0x1.head["a"] 0x7f38ec014010)
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 mds is 0
2018-07-25 20:58:04.185 7f3928b8b700 10 client.4212 send_request rebuilding request 9 for mds.0
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 encode_cap_releases enter (req: 0x1afcdd0, mds
: 0)
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 encode_dentry_release enter(dn:0x1af8960)
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 encode_inode_release enter(in:0x10000000006.he
ad(faked_ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:
58:04.185860 mtime=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsxLsXsxFsx(0
=pAsxLsXsxFsx) COMPLETE parents=0x1.head["a"] 0x7f38ec014010), req:0x1afcdd0 mds:0, drop:256, unle
ss:512, have:, force:1)
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 send_request set sent_stamp to 2018-07-25 20:5
8:04.188129
2018-07-25 20:58:04.185 7f3928b8b700 10 client.4212 send_request client_request(unknown.0:9 mkdir
 #0x10000000006/. 2018-07-25 20:58:04.188079 caller_uid=1163, caller_gid=1163{}) v4 to mds.0
2018-07-25 20:58:04.185 7f3928b8b700 20 client.4212 awaiting reply|forward|kick on 0x7ffeedc2df10
2018-07-25 20:58:04.185 7f38faffd700 10 client.4212  mds.0 seq now 3
2018-07-25 20:58:04.185 7f38faffd700  5 client.4212 handle_cap_grant on in 0x10000000006 mds.0 seq
 2 caps now pAsLsXsFsx was pAsxLsXsxFsx
2018-07-25 20:58:04.185 7f38faffd700 10 client.4212 update_inode_file_time 0x10000000006.head(fake
d_ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.1
85860 mtime=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsxLsXsxFsx(0=pAsxLs
XsxFsx) COMPLETE parents=0x1.head["a"] 0x7f38ec014010) pAsxLsXsxFsx ctime 2018-07-25 20:58:04.1858
60 mtime 2018-07-25 20:58:04.185860
2018-07-25 20:58:04.185 7f38faffd700 10 client.4212   revocation of AxXx
2018-07-25 20:58:04.185 7f38faffd700 10 client.4212 check_caps on 0x10000000006.head(faked_ino=0 r
ef=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.185860 mti
me=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsLsXsFsx(0=pAsLsXsFsx) COMPL
ETE parents=0x1.head["a"] 0x7f38ec014010) wanted Fx used - issued pAsLsXsFsx revoking AxXx flags=0
2018-07-25 20:58:04.185 7f38faffd700 10 client.4212  cap mds.0 issued pAsLsXsFsx implemented pAsxL
sXsxFsx revoking AxXx
2018-07-25 20:58:04.185 7f38faffd700 10 client.4212 completed revocation of AxXx
2018-07-25 20:58:04.185 7f38faffd700 10 client.4212 send_cap 0x10000000006.head(faked_ino=0 ref=5
 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.185860 mtime=20
18-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsLsXsFsx(0=pAsLsXsFsx) COMPLETE p
arents=0x1.head["a"] 0x7f38ec014010) mds.0 seq 2 async  used - want Fx flush - retain pAsLsxXsFsxc
rwbl held pAsxLsXsxFsx revoking AxXx dropping -
2018-07-25 20:58:04.185 7f38faffd700 15 client.4212 auth cap, setting max_size = 0
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 handle_client_reply got a reply. Safe:0 tid 9
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 insert_trace from 2018-07-25 20:58:04.188129 m
ds.0 is_target=1 is_dentry=1
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212  features 0x3ffddff8ffa4ffff
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 update_snap_trace len 48
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 get_snap_realm 0x1 0x7f38ec007da0 6 -> 7
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 update_snap_trace snaprealm(0x1 nref=7 c=0 seq
=1 parent=0x0 my_snaps=[] cached_snapc=1=[]) seq 1 <= 1 and same parent, SKIPPING
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212  hrm  is_target=1 is_dentry=1
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 update_inode_file_time 0x10000000007.head(fake
d_ino=0 ref=0 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.1
88079 mtime=0.000000 ctime=0.000000 caps=- 0x7f38ec016a10) - ctime 2018-07-25 20:58:04.188079 mtim
e 2018-07-25 20:58:04.188079
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212  dir hash is 2
2018-07-25 20:58:04.189 7f38faffd700 12 client.4212 add_update_inode adding 0x10000000007.head(fak
ed_ino=0 ref=0 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.
188079 mtime=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:58:04.188079 caps=- 0x7f38ec016a10) ca
ps pAsxLsXsxFsx
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 get_snap_realm 0x1 0x7f38ec007da0 7 -> 8
2018-07-25 20:58:04.189 7f38faffd700 15 client.4212 add_update_cap first one, opened snaprealm 0x7
f38ec007da0
```

```
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 add_update_cap issued - -> pAsxLsXsxFsx from m
ds.0 on 0x10000000007.head(faked_ino=0 ref=0 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlin
k=1 btime=2018-07-25 20:58:04.188079 mtime=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:58:04.18
8079 caps=pAsxLsXsxFsx(0=pAsxLsXsxFsx) 0x7f38ec016a10)
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212  marking (I_COMPLETE|I_DIR_ORDERED) on empty d
ir 0x10000000007.head(faked_ino=0 ref=0 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 b
time=2018-07-25 20:58:04.188079 mtime=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:58:04.188079
caps=pAsxLsXsxFsx(0=pAsxLsXsxFsx) 0x7f38ec016a10)
2018-07-25 20:58:04.189 7f38faffd700 12 client.4212 add_update_inode had 0x10000000006.head(faked_
ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.185
860 mtime=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsLsXsFsx(0=pAsLsXsFsx
) COMPLETE parents=0x1.head["a"] 0x7f38ec014010) caps pAsLsXsFsx
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 update_inode_file_time 0x10000000006.head(fake
d_ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.1
85860 mtime=2018-07-25 20:58:04.185860 ctime=2018-07-25 20:58:04.185860 caps=pAsLsXsFsx(0=pAsLsXsF
sx) COMPLETE parents=0x1.head["a"] 0x7f38ec014010) pAsLsXsFsx ctime 2018-07-25 20:58:04.188079 mti
me 2018-07-25 20:58:04.188079
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212  dir hash is 2
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 add_update_cap issued pAsLsXsFsx -> pAsLsXsFsx
 from mds.0 on 0x10000000006.head(faked_ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0
/0 nlink=1 btime=2018-07-25 20:58:04.185860 mtime=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:5
8:04.188079 caps=pAsLsXsFsx(0=pAsLsXsFsx) COMPLETE parents=0x1.head["a"] 0x7f38ec014010)
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 got dirfrag map for 0x10000000006 frag * to md
s -1
2018-07-25 20:58:04.189 7f38faffd700 12 client.4212 insert_dentry_inode '.' vino 0x10000000007.hea
d in dir 0x10000000006.head dn 0x1af8960
2018-07-25 20:58:04.189 7f38faffd700 15 inode.get on 0x7f38ec016a10 0x10000000007.head now 1
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212  clearing I_DIR_ORDERED on 0x10000000006.head(
faked_ino=0 ref=5 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:
04.185860 mtime=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:58:04.188079 caps=pAsLsXsFsx(0=pAsL
sXsFsx) COMPLETE parents=0x1.head["a"] 0x7f38ec014010)
2018-07-25 20:58:04.189 7f38faffd700 15 client.4212 link dir 0x7f38ec014010 '.' to inode 0x7f38ec0
16a10 dn 0x1af8960 (old dn)
2018-07-25 20:58:04.189 7f38faffd700 15 inode.get on 0x7f38ec016a10 0x10000000007.head now 2
2018-07-25 20:58:04.189 7f38faffd700 15 inode.get on 0x7f38ec016a10 0x10000000007.head now 3
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 put_inode on 0x10000000007.head(faked_ino=0 re
f=3 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.188079 mtim
e=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:58:04.188079 caps=pAsxLsXsxFsx(0=pAsxLsXsxFsx) CO
MPLETE parents=0x10000000006.head["."] 0x7f38ec016a10)
2018-07-25 20:58:04.189 7f38faffd700 15 inode.put on 0x7f38ec016a10 0x10000000007.head now 2
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 link  inode 0x7f38ec016a10 parents now 0x10000
000006.head["."]
2018-07-25 20:58:04.189 7f38faffd700 10 client.4212 put_inode on 0x10000000007.head(faked_ino=0 re
f=2 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.188079 mtim
e=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:58:04.188079 caps=pAsxLsXsxFsx(0=pAsxLsXsxFsx) CO
MPLETE parents=0x10000000006.head["."] 0x7f38ec016a10)
2018-07-25 20:58:04.189 7f38faffd700 15 inode.put on 0x7f38ec016a10 0x10000000007.head now 1
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 put_snap_realm 0x1 0x7f38ec007da0 8 -> 7
2018-07-25 20:58:04.189 7f38faffd700 15 inode.get on 0x7f38ec016a10 0x10000000007.head now 2
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 handle_client_reply signalling caller 0x7ffeed
c2df10
2018-07-25 20:58:04.189 7f38faffd700 20 client.4212 handle_client_reply awaiting kickback on tid 9
 0x7f38faffcbc0
2018-07-25 20:58:04.189 7f3928b8b700 20 client.4212 sendrecv kickback on tid 9 0x7f38faffcbc0
2018-07-25 20:58:04.189 7f3928b8b700 15 inode.get on 0x7f38ec016a10 0x10000000007.head now 3
2018-07-25 20:58:04.189 7f3928b8b700 10 client.4212 put_inode on 0x1.head(faked_ino=0 ref=5 ll_ref
=1 cap_refs={} open={} mode=40755 size=0/0 nlink=1 btime=2018-07-25 20:47:47.484463 mtime=2018-07-
25 20:56:49.333650 ctime=2018-07-25 20:58:04.185860 caps=pAsLsXs(0=pAsLsXs) has_dir_layout 0x7f38e
c008720)
2018-07-25 20:58:04.189 7f3928b8b700 15 inode.put on 0x7f38ec008720 0x1.head now 4
2018-07-25 20:58:04.189 7f3928b8b700 20 client.4212 make_request target is 0x10000000007.head(fake
d_ino=0 ref=3 ll_ref=0 cap_refs={} open={} mode=40775 size=0/0 nlink=1 btime=2018-07-25 20:58:04.1
88079 mtime=2018-07-25 20:58:04.188079 ctime=2018-07-25 20:58:04.188079 caps=pAsxLsXsxFsx(0=pAsxLs
XsxFsx) COMPLETE parents=0x10000000006.head["."] 0x7f38ec016a10)
2018-07-25 20:58:04.189 7f3928b8b700 20 client.4212 lat 0.003218
2018-07-25 20:58:04.189 7f3928b8b700 10 client.4212 _mkdir result is 0
2018-07-25 20:58:04.189 7f3928b8b700 20 client.4212 trim_cache size 6 max 16384
```

```
2018-07-25 20:58:04.189 7f3928b8b700  8 client.4212 _mkdir(#0x10000000006/., 040775) = 0
2018-07-25 20:58:04.189 7f3928b8b700 20 client.4212 mkdirs: successfully created directory
```

We should double check no other types of requests may create dirents with these names.

Related issues:

Copied to fs - Backport #32103: luminous: mds: allows client to create ".." a...	**Resolved**
Copied to fs - Backport #32104: mimic: mds: allows client to create ".." and ...	**Resolved**

History

#1 - 08/06/2018 01:45 PM - Patrick Donnelly

- Status changed from Verified to In Progress

#2 - 08/07/2018 01:16 PM - Venky Shankar

- Status changed from In Progress to Need Review

PR: https://github.com/ceph/ceph/pull/23469

#3 - 08/25/2018 07:56 PM - Patrick Donnelly

- Status changed from Need Review to Pending Backport

#4 - 08/28/2018 11:11 AM - Nathan Cutler

- Copied to Backport #32103: luminous: mds: allows client to create ".." and "." dirents added

#5 - 08/28/2018 11:11 AM - Nathan Cutler

- Copied to Backport #32104: mimic: mds: allows client to create ".." and "." dirents added

#6 - 10/19/2018 10:53 PM - Nathan Cutler

- Status changed from Pending Backport to Resolved

#7 - 11/19/2018 04:21 PM - Марк Коренберг

Is it possible to create such direntries using this sequense?

1. create symlink "hack" -> ".."
2. mkdir hack

i.e. create a dir by symlink. In VFS it will not work, since it first resolves symlink. I don't know how this works in Ceph (I mean mkdir/open(O_CREAT)/mknod/... on symlink pointing to missing location)

I don't follow the patch and have no ability to test before asking, sorry.