

CephFS - Bug #23289

mds: xfstest generic/089 hangs in rename syscall in luminous

03/09/2018 03:46 PM - Luis Henriques

Status:	New	% Done:	0%
Priority:	Normal		
Assignee:			
Category:			
Target version:			
Source:	Community (user)	ceph-qa-suite:	
Tags:		Component(FS):	MDS
Backport:	luminous	Labels (FS):	multimds
Regression:	No	Pull request ID:	
Severity:	3 - minor	Crash signature (v1):	
Reviewed:		Crash signature (v2):	
Affected Versions:	v12.2.4		

Description

Running this test on a multimds Luminous cluster results in a stalled (very slow?) client. I'm using a kernel client, but I was able to reproduce the issue with fuse as well. Thus, I don't think this is a client issue. Also, I wasn't able to reproduce it with a mimic (master branch) cluster.

In the client, here's the test stack:

```
cat /proc/1065/stack
[<0>] ceph_mdsc_do_request+0xef/0x2e0
[<0>] ceph_rename+0x125/0x1e0
[<0>] vfs_rename+0x335/0x810
[<0>] SyS_renameat2+0x47a/0x530
[<0>] do_syscall_64+0x60/0x110
[<0>] entry_SYSCALL_64_after_hwframe+0x3d/0xa2
[<0>] 0xffffffffffff
```

I'm able to reproduce this very easily with a vstart cluster:

```
MON=1 OSD=2 MDS=3 ../src/vstart.sh -x -n -i 192.168.155.1 --multimds 2 -b
```

If, while the test is stalled, I reduce the number of MDSs, the test will proceed and finish. I'm still looking at the MDS logs (attached), but I'm not very proficient reading those so it may take a while for me to see something that may be obvious to more trained eyes.

History

#1 - 03/10/2018 12:38 AM - Patrick Donnelly

- Subject changed from *xfstest generic/089 hangs in rename syscall in luminous* to *mds: xfstest generic/089 hangs in rename syscall in luminous*
- Target version deleted (v12.2.4)
- Source set to *Community (user)*
- Backport set to *luminous*
- Affected Versions v12.2.4 added

#2 - 03/12/2018 01:43 PM - Zheng Yan

- Assignee set to *Zheng Yan*

#3 - 03/13/2018 02:58 AM - Zheng Yan

could you check if the test is hang or slow (check `/sys/kernel/debug/ceph/xxx/mdsc`). my local test show the test is slow because there are lots of cross-mds rename

#4 - 03/13/2018 09:50 AM - Luis Henriques

I can confirm that the test still progresses (I see changes in the /sys/kernel/debug/ceph/xxx/mdsc file). But this makes the test take way too long compared to what happens in mimic (just a few minutes in a vstart cluster).

#5 - 03/14/2018 10:31 AM - Luis Henriques

By the way, have you actually seen the test run to completion? I've waited a few hours and the test kept running.

#6 - 03/09/2019 12:16 AM - Patrick Donnelly

- Assignee deleted (Zheng Yan)

#7 - 03/09/2019 12:32 AM - Patrick Donnelly

- Category deleted (90)

- Labels (FS) multimds added

Files

mds-logs.tgz	91.6 KB	03/09/2018	Luis Henriques
--------------	---------	------------	----------------