

Ceph - Bug #13662

ext4 xattr linux 3.16 / 4.2.3 panic

10/31/2015 03:35 AM - Loïc Dachary

Status:	Resolved	% Done:	0%
Priority:	Normal	Spent time:	0.00 hour
Assignee:			
Category:			
Target version:			
Source:	other	Affected Versions:	
Tags:		ceph-qa-suite:	
Backport:		Pull request ID:	
Regression:	No	Crash signature (v1):	
Severity:	3 - minor	Crash signature (v2):	
Reviewed:			
Description			
problem description			
<p>Machines hosting OSDs with Debian jessie and running linux 3.16 crashed one after the other with the attached backtrace (photo of the console). They are all related to ext4 xattr but no trace is the same.</p> <p>Running hammer 0.94.5. It looks like something related to xattr propagated to all OSD and triggered an ext4 kernel bug that has the machine crash.</p>			
workaround			
<p>Rebuilt the ubuntu trusty 3.19 kernel with the following patch: https://bugzilla.kernel.org/show_bug.cgi?id=107301#c6 The patch simply removes the mbcache code from ext4 xattr handing.</p>			
Related issues:			
Related to Ceph - Bug #14505: "kernel: BUG: soft lockup - CPU#1 stuck" in upg...		Rejected	01/25/2016

History

#1 - 10/31/2015 03:36 AM - Loïc Dachary

Mehdi & Laurent : could you add as much detail as possible about the configuration of the machines where this crash happens ?

#2 - 10/31/2015 03:37 AM - Loïc Dachary

- Subject changed from ext4 xattr linux 3.16 backtrace to ext4 xattr linux 3.16 panic

#3 - 10/31/2015 03:40 AM - Mehdi Abaakouk

- File bt_g1 added

Just got the same kind of backtrace on a 4.2.3 kernel

#4 - 10/31/2015 03:47 AM - Mehdi Abaakouk

- File bt_n7 added

Backtrace on a 3.16.7 kernel

#5 - 10/31/2015 03:55 AM - Loïc Dachary

s1

mb_cache_entry_get [mbcache]

```
ext4_xattr_block_set [ext4]  
ext4_xattr_find_entry [ext4]  
ext4_xattr_set_entry [ext4]  
ext4_xattr_set_handle [ext4]  
ext4_xattr_set [ext4]  
generic_setxattr  
__vfs_setxattr_noperm  
vfs_setxattr  
setxattr  
ext4_xattr_user_list  
ext4_xattr_list_entries  
ext4_listxattr  
__sb_start_write
```

s2

```
__mb_cache_entry_release+0x9d/0x120 [mbcache]  
ext4_xattr_get+XXXX [ext4]  
generic_getxattr  
getxattr  
ext4_discard_preallocation
```

s3

```
ext4_xattr_cache_insert [ext4]  
ext4_xattr_get [ext4]  
generic_getxattr  
vfs_getxattr  
ext4_discard_preallocations  
do_flip_open
```

#6 - 10/31/2015 04:01 AM - Loïc Dachary

- Subject changed from *ext4 xattr linux 3.16 panic* to *ext4 xattr linux 3.16 / 4.2.3 panic*

#7 - 10/31/2015 04:11 AM - Loïc Dachary

- <http://tracker.ceph.com/issues/11581#note-2> has and ext4 related panic but the problem does not look the same

#8 - 10/31/2015 06:05 AM - Laurent GUERBY

On some machines before freezing we have soft lockup messages:

```
root@g2:~#  
Message from syslogd@g2 at Oct 30 21:01:37 ...  
kernel:[76005.122564] BUG: soft lockup - CPU#3 stuck for 23s! [ceph-osd:5530]  
  
Message from syslogd@g2 at Oct 30 21:01:37 ...  
kernel:[76005.146542] BUG: soft lockup - CPU#5 stuck for 23s! [ceph-osd:5790]  
  
Message from syslogd@g2 at Oct 30 21:01:37 ...  
kernel:[76005.158531] BUG: soft lockup - CPU#6 stuck for 23s! [ceph-osd:10587]  
  
Message from syslogd@g2 at Oct 30 21:01:37 ...  
kernel:[76005.170521] BUG: soft lockup - CPU#7 stuck for 23s! [ceph-osd:5920]  
  
Message from syslogd@g2 at Oct 30 21:02:05 ...  
kernel:[76033.097887] BUG: soft lockup - CPU#3 stuck for 22s! [ceph-osd:5530]  
  
Message from syslogd@g2 at Oct 30 21:02:05 ...  
kernel:[76033.121866] BUG: soft lockup - CPU#5 stuck for 22s! [ceph-osd:5790]  
  
Message from syslogd@g2 at Oct 30 21:02:05 ...  
kernel:[76033.133855] BUG: soft lockup - CPU#6 stuck for 22s! [ceph-osd:10587]  
  
Message from syslogd@g2 at Oct 30 21:02:05 ...  
kernel:[76033.145844] BUG: soft lockup - CPU#7 stuck for 22s! [ceph-osd:5920]
```

#9 - 11/03/2015 03:16 PM - Samuel Just

We need to send a description of this to the ext4 list asap.

#10 - 11/03/2015 03:17 PM - Samuel Just

- Priority changed from *Urgent* to *High*

#11 - 11/03/2015 03:22 PM - Loïc Dachary

- Status changed from 12 to Need More Info

It would be useful to have a way to reproduce the problem. Or maybe a list of the xattr of all objects on the faulty OSD ?

#12 - 11/05/2015 01:55 PM - Mehdi Abaakouk

- File g6.dmesg added

I have opened the issue on kernel side: https://bugzilla.kernel.org/show_bug.cgi?id=107301

And got a complete backtrace via netconsole

#13 - 11/06/2015 01:22 PM - Mehdi Abaakouk

The issue seems due to a concurrency issue in the ext4 mbcache/xattr code in kernel.

In that case, does 'filestore op threads = 1' can workaround the issue ?

#14 - 11/12/2015 04:19 PM - Loïc Dachary

@laurent @mehdi : could you please summarize what you did to work around the problem ?

#15 - 11/12/2015 04:35 PM - Laurent GUERBY

Since no user level workaround were effective we rebuilt the ubuntu trusty 3.19 kernel with the following patch:

https://bugzilla.kernel.org/show_bug.cgi?id=107301#c6

The patch simply removes the mbcache code from ext4 xattr handing.

The kernel developpers are discussing what to do in the bugzilla, Sage and the Lustre developpers stepped in as users.

No issue so far in two days of production on our 11 machines with the new kernel.

#16 - 11/12/2015 06:07 PM - Loïc Dachary

- Description updated

- Status changed from Need More Info to 15

- Priority changed from High to Normal

#17 - 01/25/2016 09:56 PM - Yuri Weinstein

- Related to Bug #14505: "kernel: BUG: soft lockup - CPU#1 stuck" in upgrade:infernalis-infernalis-distro-basic-openstack added

#18 - 03/31/2016 01:56 PM - Anonymous

I just updated the korg bugzilla and updating this ticket accordingly.

The 4.6.x series of the kernel will feature a patch from Jan Kara that should solve this issue but upstream in the Kernel.

If Laurent & Mehdi can confirm that 4.6.x is safe that would be ideal to report Ceph users considering using such kernel to avoid the issue you had.

Files				
s1.png		18.1 KB	10/31/2015	Loïc Dachary
s2.png		15.4 KB	10/31/2015	Loïc Dachary
s3.png		14.9 KB	10/31/2015	Loïc Dachary
bt_g1		14.7 KB	10/31/2015	Mehdi Abaakouk
bt_n7		16.1 KB	10/31/2015	Mehdi Abaakouk
g6.dmesg		159 KB	11/05/2015	Mehdi Abaakouk