

## RADOS - Bug #22544

### objecter cannot resend split-dropped op when racing with con reset

12/27/2017 04:29 AM - mingxin liu

<b>Status:</b> Resolved	<b>Start date:</b> 12/27/2017
<b>Priority:</b> Urgent	<b>Due date:</b>
<b>Assignee:</b>	<b>% Done:</b> 0%
<b>Category:</b>	<b>Estimated time:</b> 0.00 hour
<b>Target version:</b>	<b>Spent time:</b> 0.00 hour
<b>Source:</b>	<b>Reviewed:</b>
<b>Tags:</b>	<b>Affected Versions:</b>
<b>Backport:</b> mimic,luminous	<b>ceph-qa-suite:</b>
<b>Regression:</b> No	<b>Component(RADOS):</b>
<b>Severity:</b> 3 - minor	<b>Pull request ID:</b>
<b>Description</b>	
<pre>if (split &amp;&amp; con &amp;&amp; con-&gt;has_features(CEPH_FEATUREMASK_RESEND_ON_SPLIT)) { return RECALC_OP_TARGET_NEED_RESEND; }</pre>	
resending depends on con features, if con was just reset, its feature bits is empty, letting this op sneaks. further more, if this op was resent finally after some new writes(it can happen because acting changed, con reset again..) , causing out of order.	
shall we move objecter resend logic from ms_handle_reset to ms_handle_connect?	
<b>Related issues:</b>	
Copied to RADOS - Backport #35843: mimic: objecter cannot resend split-droppe...	<b>Resolved</b>
Copied to RADOS - Backport #35844: luminous: objecter cannot resend split-dro...	<b>Resolved</b>

### History

#### #1 - 01/01/2018 10:07 AM - John Spray

- Project changed from Ceph to RADOS

#### #2 - 01/03/2018 09:09 PM - Josh Durgin

- Priority changed from Normal to Urgent

#### #3 - 01/17/2018 02:59 PM - Sage Weil

- Status changed from New to Verified

Hmm, I'm not sure what the best fix is. Do you see a good path to fixing this with ms\_handle\_connect()?

#### #4 - 09/01/2018 08:43 PM - Sage Weil

Here, it happened:

```
2018-08-31 20:50:46.286 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 155 ===== osd_op(client.4338.0:9206 2.5s0 2.60bf6c05 (undecoded) ondisk+write+known_if_redirected e84) v8 =
=== 526+0+114688 (2347694735 0 2580345875) 0x558626969a00 con 0x558626eeb100
2018-08-31 20:50:46.286 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 156 ===== osd_op(client.4338.0:9207 2.5s0 2.60bf6c05 (undecoded) ondisk+write+known_if_redirected e84) v8 =
=== 526+0+363 (2433957676 0 3325077087) 0x558626968000 con 0x558626eeb100
2018-08-31 20:50:46.286 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 157 ===== osd_op(client.4338.0:9208 2.5s0 2.60bf6c05 (undecoded) ondisk+read+rwordered+known_if_redirected
e84) v8 ===== 526+0+0 (1514364427 0 0) 0x558625722080 con 0x558626eeb100
```

```
2018-08-31 20:50:46.342 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 1 ==== osd_op(client.4338.0:9226 2.5s0 2.9d089415 (undecoded) ondisk+retry+write+known_if_redirected e84)
v8 ==== 263+0+614400 (405944600 0 301684070) 0x558627126080 con 0x5586261f8700
2018-08-31 20:50:46.342 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 2 ==== osd_op(client.4338.0:9227 2.15s0 2.9d089415 (undecoded) ondisk+write+known_if_redirected e85) v8 ==
== 225+0+442368 (4173242535 0 1877614199) 0x558626cdda00 con 0x5586261f8700
2018-08-31 20:50:46.342 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 3 ==== osd_op(client.4338.0:9228 2.15s0 2.9d089415 (undecoded) ondisk+write+known_if_redirected e85) v8 ==
== 225+0+360448 (2218567718 0 4227264512) 0x5586277d1040 con 0x5586261f8700
2018-08-31 20:50:46.342 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 4 ==== osd_op(client.4338.0:9229 2.15s0 2.9d089415 (undecoded) ondisk+write+known_if_redirected e85) v8 ==
== 225+0+62 (3421199549 0 11496458) 0x5586271ec340 con 0x5586261f8700
2018-08-31 20:50:46.342 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 5 ==== osd_op(client.4338.0:9230 2.15s0 2.9d089415 (undecoded) ondisk+read+rworderred+known_if_redirected e
85) v8 ==== 225+0+0 (118536143 0 0) 0x5586271ecc00 con 0x5586261f8700
2018-08-31 20:50:48.514 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 6 ==== osd_op(client.4338.0:9247 2.bs0 2.9faadfbb (undecoded) ondisk+write+known_if_redirected e86) v8 ==
= 527+0+770048 (4084052600 0 914887116) 0x558627ff2680 con 0x5586261f8700
2018-08-31 20:50:48.518 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 7 ==== osd_op(client.4338.0:9248 2.bs0 2.9faadfbb (undecoded) ondisk+write+known_if_redirected e86) v8 ==
= 527+0+655360 (3769274892 0 2961804097) 0x558627ff29c0 con 0x5586261f8700
2018-08-31 20:50:48.530 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 1 ==== osd_op(client.4338.0:9226 2.15s0 2.9d089415 (undecoded) ondisk+retry+write+known_if_redirected e86)
v8 ==== 263+0+614400 (3320940746 0 301684070) 0x558626f24680 con 0x558626f99500
2018-08-31 20:50:48.530 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 2 ==== osd_op(client.4338.0:9247 2.bs0 2.9faadfbb (undecoded) ondisk+retry+write+known_if_redirected e86)
v8 ==== 527+0+770048 (343352739 0 914887116) 0x558627ff2d00 con 0x558626f99500
2018-08-31 20:50:48.534 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 3 ==== osd_op(client.4338.0:9248 2.bs0 2.9faadfbb (undecoded) ondisk+retry+write+known_if_redirected e86)
v8 ==== 527+0+655360 (128976343 0 2961804097) 0x558625eb4a40 con 0x558626f99500
2018-08-31 20:50:48.534 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 4 ==== osd_op(client.4338.0:9249 2.bs0 2.9faadfbb (undecoded) ondisk+write+known_if_redirected e86) v8 ==
= 527+0+147456 (2079555912 0 2676445555) 0x558627ff36c0 con 0x558626f99500
2018-08-31 20:50:48.534 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 5 ==== osd_op(client.4338.0:9250 2.bs0 2.9faadfbb (undecoded) ondisk+write+known_if_redirected e86) v8 ==
= 527+0+364 (566634381 0 2994847827) 0x558627ff4080 con 0x558626f99500
2018-08-31 20:50:48.534 7fa1d0e6e700 1 -- 172.21.15.135:6800/10478 <== client.4338 172.21.15.62:58820/2203252
198 6 ==== osd_op(client.4338.0:9251 2.bs0 2.9faadfbb (undecoded) ondisk+read+rworderred+known_if_redirected e8
6) v8 ==== 527+0+0 (4294433175 0 0) 0x558627ff4a40 con 0x558626f99500
```

notice 9226 and 9227 pgids and osd epochs.

/a/sage-2018-08-31\_18:31:48-rados-wip-sage-testing-2018-08-31-1010-distro-basic-smithi/2964779

**#5 - 09/01/2018 08:43 PM - Sage Weil**

- Backport set to *mimic,luminous*

**#6 - 09/01/2018 08:49 PM - Sage Weil**

- Status changed from *Verified* to *Need Review*

<https://github.com/ceph/ceph/pull/23850>

**#7 - 09/06/2018 12:33 AM - Kefu Chai**

- Status changed from *Need Review* to *Pending Backport*

**#8 - 09/07/2018 10:22 AM - Nathan Cutler**

- Copied to Backport #35843: *mimic: objecter cannot resend split-dropped op when racing with con reset added*

**#9 - 09/07/2018 10:22 AM - Nathan Cutler**

- Copied to Backport #35844: *luminous: objecter cannot resend split-dropped op when racing with con reset added*

**#10 - 11/26/2018 11:48 PM - Nathan Cutler**

- Status changed from *Pending Backport* to *Resolved*