fs - Bug #38203

ceph-mds segfault during migrator nicely exporting

02/06/2019 10:34 AM - Dan van der Ster

Status:	New	**Start date:**	
Priority:	Urgent	**Due date:**	
Assignee:	Patrick Donnelly	**% Done:**	0%
Category:		**Estimated time:**	0.00 hour
Target version:	v15.0.0		
Source:	Community (user)	**Affected Versions:**	v12.2.10
Tags:		**ceph-qa-suite:**	
Backport:	mimic,luminous	**Component(FS):**	MDS
Regression:	No	**Labels (FS):**	multimds
Severity:	3 - minor	**Pull request ID:**	
Reviewed:			

Description

We had a v12.2.10 mds crash during md balancing:

```
   -22> 2019-02-06 03:49:51.932215 7f7be2bc9700  0 mds.3.migrator nicely exporting to mds.0 [dir 0
x500006410dd /volumes/_nogroup/69d07521-b740-46d1-a545-7057160a8fb2/user/atlact1/www/jobs/2019-02-
04/ [2,head] auth{0=1,1=1,2=1,4=1} v=744446 cv=744381/744381 REP dir_auth=3 state=1610612738|compl
ete f(v1 m2019-02-05 10:16:58.337992 57=0+57) n(v9947 rc2019-02-06 03:49:44.347839 b110305046691 1
80971=180914+57)/n(v9947 rc2019-02-06 03:49:37.147924 b110298302472 180966=180909+57) hs=57+0,ss=0
+0 dirty=1 | dnwaiter=0 child=1 frozen=0 subtree=1 importing=0 replicated=1 dirty=1 waiter=0 authp
in=0 0x561f5f294e00]
   -21> 2019-02-06 03:49:52.583911 7f7be63d0700  2 mds.3.cache check_memory_usage total 14002256,
rss 11375468, heap 315424, baseline 315424, buffers 0, 55325 / 2141345 inodes have caps, 83015 cap
s, 0.0387677 caps per inode
   -20> 2019-02-06 03:49:54.692285 7f7be5bcf700 10 monclient: _send_mon_message to mon.cephflax-mo
n-9b406e0261 at 137.138.121.135:6789/0
   -19> 2019-02-06 03:49:56.929179 7f7be7bd3700 10 monclient: tick
   -18> 2019-02-06 03:49:56.929232 7f7be7bd3700 10 monclient: _check_auth_rotating have uptodate s
ecrets (they expire after 2019-02-06 03:49:26.929206)
   -17> 2019-02-06 03:49:57.635358 7f7be63d0700  2 mds.3.cache check_memory_usage total 14002256,
rss 11375172, heap 315424, baseline 315424, buffers 0, 55334 / 2133234 inodes have caps, 83029 cap
s, 0.0389217 caps per inode
   -16> 2019-02-06 03:49:58.692466 7f7be5bcf700 10 monclient: _send_mon_message to mon.cephflax-mo
n-9b406e0261 at 137.138.121.135:6789/0
   -15> 2019-02-06 03:50:01.357112 7f7be2bc9700  0 mds.3.migrator nicely exporting to mds.0 [dir 0
x100093748b9.101* /volumes/_nogroup/f51e1a94-b1ff-4f2c-9686-c7e9835e0f6c/work/cli-build/ [2,head]
auth{0=1} v=437725 cv=437725/437725 REP dir_auth=3 state=1073741826|complete f(v48 m2019-02-05 15:
22:45.539666 8788=0+8788) n(v5346 rc2019-02-05 15:23:10.481882 b105348841048 24087=15299+8788) hs=
8788+0,ss=0+0 | dnwaiter=0 child=1 frozen=0 subtree=1 importing=0 replicated=1 dirty=0 waiter=0 au
thpin=0 tempexporting=0 0x561f77cfa000]
   -14> 2019-02-06 03:50:01.523391 7f7be8bd5700  0 mds.3.migrator nicely exporting to mds.0 [dir 0
x100093748b9.010* /volumes/_nogroup/f51e1a94-b1ff-4f2c-9686-c7e9835e0f6c/work/cli-build/ [2,head]
auth{0=1,2=1} v=426986 cv=426986/426986 REP dir_auth=3 state=1073741826|complete f(v48 m2019-02-06
 00:38:56.760195 8769=0+8769) n(v5346 rc2019-02-06 00:39:12.952094 b100212238080 24118=15349+8769)
 hs=8769+0,ss=0+0 | dnwaiter=0 child=1 frozen=0 subtree=1 importing=0 replicated=1 0x561f875d1800]
   -13> 2019-02-06 03:50:02.190203 7f7be93d6700  5 asok(0x561da983e1c0) AdminSocket: request 'get_
command_descriptions' '' to 0x561da98201e0 returned 4774 bytes
   -12> 2019-02-06 03:50:02.208436 7f7be93d6700  1 do_command 'perf dump' '
   -11> 2019-02-06 03:50:02.208745 7f7be93d6700  1 do_command 'perf dump' 'result is 13156 bytes
   -10> 2019-02-06 03:50:02.208755 7f7be93d6700  5 asok(0x561da983e1c0) AdminSocket: request 'perf
 dump' '' to 0x561da9820040 returned 13156 bytes
    -9> 2019-02-06 03:50:02.209381 7f7be93d6700  5 asok(0x561da983e1c0) AdminSocket: request 'get_
command_descriptions' '' to 0x561da98201e0 returned 4774 bytes
    -8> 2019-02-06 03:50:02.227158 7f7be93d6700  1 mds.cephflax-mds-2a4cfd0e2c asok_command: sessi
```

```
on ls (starting...)
    -7> 2019-02-06 03:50:02.234906 7f7be93d6700  1 mds.cephflax-mds-2a4cfd0e2c asok_command: sessi
on ls (complete)
    -6> 2019-02-06 03:50:02.236787 7f7be93d6700  5 asok(0x561da983e1c0) AdminSocket: request 'sess
ion ls' '' to 0x561da9820220 returned 420758 bytes
    -5> 2019-02-06 03:50:02.302902 7f7beb355700  0 -- 137.138.52.159:6800/1556042223 >> 10.32.3.22
3:0/1782848130 conn(0x561dd03b7800 :6800 s=STATE_OPEN pgs=1008711 cs=1 l=0).process bad tag 50
    -4> 2019-02-06 03:50:02.569959 7f7beb355700  0 -- 137.138.52.159:6800/1556042223 >> 10.32.3.22
3:0/1782848130 conn(0x561dbecb3800 :6800 s=STATE_ACCEPTING_WAIT_CONNECT_MSG_AUTH pgs=0 cs=0 l=0).h
andle_connect_msg accept connect_seq 2 vs existing csq=1 existing_state=STATE_STANDBY
    -3> 2019-02-06 03:50:02.641677 7f7be63d0700  2 mds.3.cache check_memory_usage total 14002256,
rss 11374236, heap 315424, baseline 315424, buffers 0, 55459 / 2107149 inodes have caps, 83164 cap
s, 0.0394675 caps per inode
    -2> 2019-02-06 03:50:02.692685 7f7be5bcf700 10 monclient: _send_mon_message to mon.cephflax-mo
n-9b406e0261 at 137.138.121.135:6789/0
    -1> 2019-02-06 03:50:04.494387 7f7beb355700  0 -- 137.138.52.159:6800/1556042223 >> 188.184.64
.209:0/3134808329 conn(0x561dd0419000 :6800 s=STATE_OPEN pgs=1128257 cs=1 l=0).process bad tag 50
     0> 2019-02-06 03:50:05.914673 7f7be8bd5700 -1 *** Caught signal (Segmentation fault) **
 in thread 7f7be8bd5700 thread_name:ms_dispatch

 ceph version 12.2.10 (177915764b752804194937482a39e95e0ca3de94) luminous (stable)
 1: (()+0x5ca6c1) [0x561d9f4036c1]
 2: (()+0xf5d0) [0x7f7beda135d0]
 3: (tcmalloc::CentralFreeList::ReleaseListToSpans(void*)+0x13) [0x7f7bef2c0cb3]
 4: (tcmalloc::CentralFreeList::ShrinkCache(int, bool)+0x124) [0x7f7bef2c0df4]
 5: (tcmalloc::CentralFreeList::MakeCacheSpace()+0x5d) [0x7f7bef2c0f1d]
 6: (tcmalloc::CentralFreeList::InsertRange(void*, void*, int)+0x68) [0x7f7bef2c0f98]
 7: (tcmalloc::ThreadCache::ReleaseToCentralCache(tcmalloc::ThreadCache::FreeList*, unsigned int,
int)+0xb8) [0x7f7bef2c4408]
 8: (tcmalloc::ThreadCache::Scavenge()+0x48) [0x7f7bef2c4798]
 9: (MDRequestImpl::~MDRequestImpl()+0x224) [0x561d9f1b5b64]
 10: (MDRequestImpl::~MDRequestImpl()+0x9) [0x561d9f1b5ce9]
 11: (OpTracker::unregister_inflight_op(TrackedOp*)+0xf2) [0x561d9f3a88e2]
 12: (Migrator::handle_export_discover_ack(MExportDirDiscoverAck*)+0x2e6) [0x561d9f2c9196]
 13: (Migrator::dispatch(Message*)+0xbb) [0x561d9f2d370b]
 14: (MDSRank::handle_deferrable_message(Message*)+0x613) [0x561d9f1079d3]
 15: (MDSRank::_dispatch(Message*, bool)+0x1e3) [0x561d9f115e13]
 16: (MDSRankDispatcher::ms_dispatch(Message*)+0x15) [0x561d9f116d55]
 17: (MDSDaemon::ms_dispatch(Message*)+0xf3) [0x561d9f0ff093]
 18: (DispatchQueue::entry()+0x792) [0x561d9f73ba22]
 19: (DispatchQueue::DispatchThread::entry()+0xd) [0x561d9f4cc9cd]
 20: (()+0x7dd5) [0x7f7beda0bdd5]
 21: (clone()+0x6d) [0x7f7becae8ead]
 NOTE: a copy of the executable, or `objdump -rdS <executable>` is needed to interpret this.
```

Log is in ceph-post-file: 3933388c-0fe6-4095-a122-11d4a4c4c6e8

Core dump is available but 14GB so it might be more convenient if you tell me what to grab from gdb.

History

#1 - 02/06/2019 03:32 PM - Patrick Donnelly

- Assignee set to Patrick Donnelly

- Priority changed from Normal to Urgent

- Target version set to v14.0.0

- Start date deleted (02/06/2019)

- Backport set to mimic,luminous

- Affected Versions v12.2.10 added

#2 - 02/12/2019 03:41 AM - Zheng Yan

please gdb the coredump, check where is MDRequestImpl::~MDRequestImpl()+0x224

#3 - 02/12/2019 09:21 AM - Dan van der Ster

```
(gdb) up
#12 0x0000561d9f1b5b64 in MDRequestImpl::~MDRequestImpl (this=0x561fdc42b700, __in_chrg=<optimized out>) at /u
sr/src/debug/ceph-12.2.10/src/mds/Mutation.cc:189
189       delete _more;
(gdb) list
184     {
185       if (client_request)
186         client_request->put();
187       if (slave_request)
188         slave_request->put();
189       delete _more;
190     }
191
192     MDRequestImpl::More* MDRequestImpl::more()
193     {
(gdb) p _more
$1 = (MDRequestImpl::More *) 0x561df4e52000
(gdb)
```

#4 - 02/12/2019 02:26 PM - Dan van der Ster

up a bit higher:

```
#14 0x0000561d9f3a88e2 in put (this=<optimized out>) at /usr/src/debug/ceph-12.2.10/src/common/TrackedOp.h:236
236         delete this;
(gdb) list
231       void put() {
232         if (--nref == 0) {
233           switch (state.load()) {
234           case STATE_UNTRACKED:
235         _unregistered();
236         delete this;
237         break;
238
239           case STATE_LIVE:
240         mark_event("done");
(gdb)
(gdb) up
#15 intrusive_ptr_release (o=<optimized out>) at /usr/src/debug/ceph-12.2.10/src/common/TrackedOp.h:315
315         o->put();
(gdb) list
310       // put for historical op tracking
311       friend void intrusive_ptr_add_ref(TrackedOp *o) {
312         o->get();
313       }
314       friend void intrusive_ptr_release(TrackedOp *o) {
315         o->put();
```

```
316       }
317     };
318
319
(gdb) up
#16 ~intrusive_ptr (this=<optimized out>, __in_chrg=<optimized out>) at /usr/src/debug/ceph-12.2.10/build/boos
t/include/boost/smart_ptr/intrusive_ptr.hpp:98
98            if( px != 0 ) intrusive_ptr_release( px );
(gdb) up
#17 OpTracker::unregister_inflight_op (this=0x561da9e172e8, i=<optimized out>) at /usr/src/debug/ceph-12.2.10/
src/common/TrackedOp.cc:283
283         history.insert(now, TrackedOpRef(i));
(gdb) list
278       if (!tracking_enabled)
279         delete i;
280       else {
281         i->state = TrackedOp::STATE_HISTORY;
282         utime_t now = ceph_clock_now();
283         history.insert(now, TrackedOpRef(i));
284       }
285     }
286
287     bool OpTracker::check_ops_in_flight(std::vector<string> &warning_vector, int *slow)
(gdb)
(gdb) up
#18 0x0000561d9f2c9196 in intrusive_ptr_release (o=<optimized out>) at /usr/src/debug/ceph-12.2.10/src/common/
TrackedOp.h:315
315         o->put();
(gdb) up
#19 ~intrusive_ptr (this=0x7f7be8bd2f10, __in_chrg=<optimized out>) at /usr/src/debug/ceph-12.2.10/build/boost
/include/boost/smart_ptr/intrusive_ptr.hpp:98
98            if( px != 0 ) intrusive_ptr_release( px );
(gdb) up
#20 Migrator::handle_export_discover_ack (this=this@entry=0x561da9e36630, m=m@entry=0x5620ade63e00) at /usr/sr
c/debug/ceph-12.2.10/src/mds/Migrator.cc:1320
1320          assert(g_conf->mds_kill_export_at != 3);
(gdb) up
#21 0x0000561d9f2d370b in Migrator::dispatch (this=0x561da9e36630, m=m@entry=0x5620ade63e00) at /usr/src/debug
/ceph-12.2.10/src/mds/Migrator.cc:146
146         handle_export_discover_ack(static_cast<MExportDirDiscoverAck*>(m));
```

#5 - 02/13/2019 08:54 AM - Zheng Yan

In MDRequestImpl::~MDRequestImpl, please check if 'state' is STATE_UNTRACKED.

In Migrator::handle_export_discover_ack, please check if 'mdr.px' is 0x561fdc42b700. (if mdr is optimized out, go down a level to ~intrusive_ptr, check px)

#6 - 02/13/2019 09:43 AM - Dan van der Ster

```
(gdb) f 12
#12 0x0000561d9f1b5b64 in MDRequestImpl::~MDRequestImpl (this=0x561fdc42b700, __in_chrg=<optimized out>)
    at /usr/src/debug/ceph-12.2.10/src/mds/Mutation.cc:189
189      delete _more;
(gdb) p state
$1 = {<std::__atomic_base<int>> = {_M_i = 2}, <No data fields>}
(gdb) f 20
#20 Migrator::handle_export_discover_ack (this=this@entry=0x561da9e36630, m=m@entry=0x5620ade63e00)
    at /usr/src/debug/ceph-12.2.10/src/mds/Migrator.cc:1320
1320          assert(g_conf->mds_kill_export_at != 3);
(gdb) p mdr.px
$2 = (MDRequestImpl *) 0x561fdc42b700
(gdb)
```

#7 - 02/13/2019 01:09 PM - Zheng Yan

Thanks. But everything looks fine. I have no idea what happened

#8 - 02/13/2019 01:13 PM - Dan van der Ster

OK, we only had this type of crash once or twice, so feel free to drop the priority.
We'll let you know if it reoccurs.

#9 - 03/07/2019 11:22 PM - Patrick Donnelly

- Target version changed from v14.0.0 to v15.0.0

#10 - 03/09/2019 12:32 AM - Patrick Donnelly

- Category deleted (90)