rgw - Bug #22084

Swift object expiry incorrectly trims entries, leaving behind some of the objects to be not deleted

11/08/2017 04:01 PM - Pavan Rallabhandi

Status:	Resolved	**% Done:**	0%
Priority:	Normal	**Spent time:**	0.00 hour
Assignee:	Pavan Rallabhandi		
Category:			
Target version:			
Source:	Community (dev)	**Affected Versions:**	
Tags:		**ceph-qa-suite:**	
Backport:	jewel luminous	**Pull request ID:**	
Regression:	No	**Crash signature (v1):**	
Severity:	3 - minor	**Crash signature (v2):**	
Reviewed:			

Description

In cls_timeindex_list() though `to_index` has expired for a timespan, the marker is set for a subsequent index during the time boundary check. This marker is further returned to RGWObjectExpirer::process_single_shard(), where this out_marker is trimmed from the respective shard, resulting in a lost removal hint and a leaked object.

To reproduce this, I've used the following simple script and by having an aggressive object expiry gc interval setting rgw_objexp_gc_interval = 30

```
!/bin/sh
object="myobjects"
container=newcontainer
objects=50
j=1
while [ $j -lt $objects ]
do
final_object=$(printf "%s$j" "$object")
touch $final_object
swift -A http://localhost:80/auth -U test:tester -K testing upload $container $final_object -H X-D
elete-After:60
j=`expr $j + 1`
done
```

For one such run, am left with the following leaked objects after the run

```
swift -A http://localhost:80/auth -U test:tester -K testing list newcontainer
myobjects20
myobjects24
myobjects33
```

Here is a snippet from RGW and OSD logs, where one can see how the delete hint for `myobjects20` is lost, find some of my extra debugs that are added

RGW logs

```
2017-11-08 15:10:30.236 7ff0f8bb7700 20 proceeding shard = obj_delete_at_hint.0000000112
2017-11-08 15:10:30.388 7ff0f8bb7700 15 got removal hint for: 1510133953 - newcontainer:7870b940-a
```

```
af0-443b-a6a3-da4f40ce9a65.4122.1:myobjects1:
2017-11-08 15:10:30.388 7ff0f8bb7700 20 reading from default.rgw.meta:root:.bucket.meta.newcontain
er:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1
2017-11-08 15:10:30.388 7ff0f8bb7700 20 get_system_obj_state: rctx=0x7ff0f8bb25e0 obj=default.rgw.
meta:root:.bucket.meta.newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 state=0x7ff11548b9
20 s->prefetch_data=0
2017-11-08 15:10:30.388 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x16, cached=0x17)
2017-11-08 15:10:30.388 7ff0f8bb7700 20 get_system_obj_state: s->obj_tag was set empty
2017-11-08 15:10:30.388 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x11, cached=0x17)
2017-11-08 15:10:30.388 7ff0f8bb7700 20 get_obj_state: rctx=0x7ff0f8bb2a50 obj=newcontainer:myobje
cts1 state=0x7ff115d5c3a0 s->prefetch_data=0
2017-11-08 15:10:30.388 7ff0f8bb7700 10 manifest: total_size = 0
2017-11-08 15:10:30.388 7ff0f8bb7700 20 get_obj_state: setting s->obj_tag to 7870b940-aaf0-443b-a6
a3-da4f40ce9a65.4122.5
2017-11-08 15:10:30.388 7ff0f8bb7700 20 get_obj_state: rctx=0x7ff0f8bb2a50 obj=newcontainer:myobje
cts1 state=0x7ff115d5c3a0 s->prefetch_data=0
2017-11-08 15:10:30.388 7ff0f8bb7700 20 reading from default.rgw.meta:root:.bucket.meta.newcontain
er:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1
2017-11-08 15:10:30.388 7ff0f8bb7700 20 get_system_obj_state: rctx=0x7ff0f8bb1a90 obj=default.rgw.
meta:root:.bucket.meta.newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 state=0x7ff11548b7
80 s->prefetch_data=0
2017-11-08 15:10:30.388 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x16, cached=0x17)
2017-11-08 15:10:30.388 7ff0f8bb7700 20 get_system_obj_state: s->obj_tag was set empty
2017-11-08 15:10:30.388 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x11, cached=0x17)
2017-11-08 15:10:30.388 7ff0f8bb7700 20  bucket index object: .dir.7870b940-aaf0-443b-a6a3-da4f40c
e9a65.4122.1
2017-11-08 15:10:30.772 7ff0f8bb7700 15 got removal hint for: 1510133966 - newcontainer:7870b940-a
af0-443b-a6a3-da4f40ce9a65.4122.1:myobjects6:
2017-11-08 15:10:30.772 7ff0f8bb7700 20 reading from default.rgw.meta:root:.bucket.meta.newcontain
er:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1
2017-11-08 15:10:30.772 7ff0f8bb7700 20 get_system_obj_state: rctx=0x7ff0f8bb25e0 obj=default.rgw.
meta:root:.bucket.meta.newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 state=0x7ff11548b7
80 s->prefetch_data=0
2017-11-08 15:10:30.772 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x16, cached=0x17)
2017-11-08 15:10:30.772 7ff0f8bb7700 20 get_system_obj_state: s->obj_tag was set empty
2017-11-08 15:10:30.772 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x11, cached=0x17)
2017-11-08 15:10:30.772 7ff0f8bb7700 20 get_obj_state: rctx=0x7ff0f8bb2a50 obj=newcontainer:myobje
cts6 state=0x7ff115d5c3a0 s->prefetch_data=0
2017-11-08 15:10:30.776 7ff0f8bb7700 10 manifest: total_size = 0
2017-11-08 15:10:30.776 7ff0f8bb7700 20 get_obj_state: setting s->obj_tag to 7870b940-aaf0-443b-a6
a3-da4f40ce9a65.4122.30
2017-11-08 15:10:30.776 7ff0f8bb7700 20 get_obj_state: rctx=0x7ff0f8bb2a50 obj=newcontainer:myobje
cts6 state=0x7ff115d5c3a0 s->prefetch_data=0
2017-11-08 15:10:30.776 7ff0f8bb7700 20 reading from default.rgw.meta:root:.bucket.meta.newcontain
er:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1
2017-11-08 15:10:30.776 7ff0f8bb7700 20 get_system_obj_state: rctx=0x7ff0f8bb1a90 obj=default.rgw.
meta:root:.bucket.meta.newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 state=0x7ff11548b7
80 s->prefetch_data=0
2017-11-08 15:10:30.776 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x16, cached=0x17)
2017-11-08 15:10:30.776 7ff0f8bb7700 20 get_system_obj_state: s->obj_tag was set empty
2017-11-08 15:10:30.776 7ff0f8bb7700 10 cache get: name=default.rgw.meta+root+.bucket.meta.newcont
ainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1 : hit (requested=0x11, cached=0x17)
2017-11-08 15:10:30.776 7ff0f8bb7700 20  bucket index object: .dir.7870b940-aaf0-443b-a6a3-da4f40c
e9a65.4122.1
2017-11-08 15:10:31.660 7ff0f8bb7700 10 out marker is 1_1510134003.861485_newcontainer:7870b940-aa
f0-443b-a6a3-da4f40ce9a65.4122.1:myobjects20:
2017-11-08 15:10:31.660 7ff0f8bb7700 20 trying to trim removal hints to=2017-11-08 15:09:33.726532
, to_marker=1_1510134003.861485_newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1:myobjects
20:
2017-11-08 15:10:32.172 7ff0f8bb7700 20 proceeding shard = obj_delete_at_hint.0000000113
```

OSD logs

```
2017-11-08 15:10:30.388 7f712e306700 20 <cls> ceph/src/cls/timeindex/cls_timeindex.cc:156: DEBUG:
cls_timeindex_list: index=1_1510133953.764091_newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.41
22.1:myobjects1:, key_ext=newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1:myobjects1:, bl
.len = 109
2017-11-08 15:10:30.388 7f712e306700 20 <cls> ceph/src/cls/timeindex/cls_timeindex.cc:159: Marker
is 1_1510133953.764091_newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1:myobjects1:
2017-11-08 15:10:30.388 7f712e306700 20 <cls> ceph/src/cls/timeindex/cls_timeindex.cc:156: DEBUG:
cls_timeindex_list: index=1_1510133966.796701_newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.41
22.1:myobjects6:, key_ext=newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1:myobjects6:, bl
.len = 109
2017-11-08 15:10:30.388 7f712e306700 20 <cls> ceph/src/cls/timeindex/cls_timeindex.cc:159: Marker
is 1_1510133966.796701_newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1:myobjects6:
2017-11-08 15:10:30.388 7f712e306700 20 <cls> ceph/src/cls/timeindex/cls_timeindex.cc:143: DEBUG:
cls_timeindex_list: finishing on to_index=1_1510133973.726532_
2017-11-08 15:10:30.388 7f712e306700 20 <cls> ceph/src/cls/timeindex/cls_timeindex.cc:145: Marker
is 1_1510134003.861485_newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1:myobjects20:
2017-11-08 15:10:30.388 7f712e306700 20 <cls> ceph/src/cls/timeindex/cls_timeindex.cc:163: Returni
ng marker as 1_1510134003.861485_newcontainer:7870b940-aaf0-443b-a6a3-da4f40ce9a65.4122.1:myobject
s20:
```

Related issues:

Copied to rgw - Backport #22179: luminous: Swift object expiry incorrectly tr...	**Resolved**
Copied to rgw - Backport #22180: jewel: Swift object expiry incorrectly trims...	**Resolved**

History

#1 - 11/08/2017 04:02 PM - Pavan Rallabhandi

- Description updated

#2 - 11/08/2017 04:52 PM - Pavan Rallabhandi

- Status changed from New to Fix Under Review

https://github.com/ceph/ceph/pull/18821

#3 - 11/08/2017 04:59 PM - Pavan Rallabhandi

- Description updated

https://github.com/ceph/ceph/pull/18821

#4 - 11/08/2017 06:10 PM - Casey Bodley

- Backport set to jewel luminous

#5 - 11/09/2017 04:37 PM - Casey Bodley

- Status changed from Fix Under Review to 7

#6 - 11/10/2017 04:04 PM - Yuri Weinstein

Pavan Rallabhandi wrote:

> https://github.com/ceph/ceph/pull/18821

merged

#7 - 11/13/2017 05:36 PM - Ken Dreyer

- Status changed from 7 to Pending Backport

#8 - 11/14/2017 03:39 PM - Casey Bodley

jewel backport pr: https://github.com/ceph/ceph/pull/18925

#9 - 11/20/2017 11:05 AM - Nathan Cutler

- Copied to Backport #22179: luminous: Swift object expiry incorrectly trims entries, leaving behind some of the objects to be not deleted added

#10 - 11/20/2017 11:05 AM - Nathan Cutler

- Copied to Backport #22180: jewel: Swift object expiry incorrectly trims entries, leaving behind some of the objects to be not deleted added

#11 - 01/23/2018 02:36 PM - Nathan Cutler

- Status changed from Pending Backport to Resolved

#12 - 01/29/2018 04:10 PM - Yuri Weinstein

https://github.com/ceph/ceph/pull/18972 merged