# Ceph - Bug #2796

## osd: watch state not reestablished when registration op resent

07/17/2012 09:04 AM - Sage Weil

| | | | |
|---|---|---|---|
| **Status:** | Resolved | **Start date:** | 07/17/2012 |
| **Priority:** | Urgent | **Due date:** | |
| **Assignee:** | | **% Done:** | 0% |
| **Category:** | Objecter | **Estimated time:** | 0.00 hour |
| **Target version:** | v0.49 | **Spent time:** | 0.00 hour |
| **Source:** | Development | **Reviewed:** | |
| **Tags:** | | **Affected Versions:** | |
| **Backport:** | argonaut | **ceph-qa-suite:** | |
| **Regression:** | No | **Pull request ID:** | |
| **Severity:** | 3 - minor | **Crash signature:** | |

**Description**

if the client doesn't get the watch ack and resends, the osd will ignore it as a dup op, and the watch session state is not reestablished.

### Associated revisions

**Revision 5dd68b95 - 07/18/2012 07:55 PM - Sage Weil**

objecter: always resend linger registrations

If a linger op (watch) is sent to the OSD and updates the object, and then the client loses the reply, it will resend the request. The OSD will see that it is a dup, however, and not set up the in-memory session state for the watch. This in turn will break the watch (i.e., notifies won't get delivered).

Instead, always resend linger registration ops, so that we always have a unique reqid and do the correct session registeration for each session.

- track the tid of the registation op for each LingerOp
- mark registrations ops as should_resend=false; cancel as needed
- when we send a new registration op, cancel the old one to ensure we ignore the reply. This is needed becuase we resend linger ops on any pg change, not just a primary change.
- drop the first_send arg to send_linger(), as we can now infer that from register_tid == 0.

The bug was easily reproduced with ms inject socket failures = 500 and the test_stress_watch utility.

Fixes: #2796
Signed-off-by: Sage Weil <sage@inktank.com>
Reviewed-by: Josh Durgin <josh.durgin@inktank.com>

**Revision 682609a9 - 07/26/2012 10:03 PM - Sage Weil**

objecter: always resend linger registrations

If a linger op (watch) is sent to the OSD and updates the object, and then the client loses the reply, it will resend the request. The OSD will see that it is a dup, however, and not set up the in-memory session state for the watch. This in turn will break the watch (i.e., notifies won't get delivered).

Instead, always resend linger registration ops, so that we always have a unique reqid and do the correct session registeration for each session.

- track the tid of the registation op for each LingerOp
- mark registrations ops as should_resend=false; cancel as needed
- when we send a new registration op, cancel the old one to ensure we ignore the reply. This is needed becuase we resend linger ops on any pg change, not just a primary change.
- drop the first_send arg to send_linger(), as we can now infer that from register_tid == 0.

The bug was easily reproduced with ms inject socket failures = 500 and the test_stress_watch utility.

Fixes: #2796
Signed-off-by: Sage Weil <sage@inktank.com>
Reviewed-by: Josh Durgin <josh.durgin@inktank.com>

## History

**#1 - 07/17/2012 12:45 PM - Sage Weil**

*- Status changed from New to Need Review*

*- Assignee deleted (Sage Weil)*

**#2 - 07/17/2012 12:45 PM - Sage Weil**

*- Backport set to argonaut*

**#3 - 07/17/2012 12:46 PM - Sage Weil**

*- Target version set to v0.49*

**#4 - 07/17/2012 07:11 PM - Sage Weil**

*- Status changed from Need Review to Testing*

**#5 - 07/18/2012 12:55 PM - Sage Weil**

*- Status changed from Testing to Resolved*