

## rgw - Bug #17822

### multisite requests failing with '400 Bad Request' with civetweb 1.8

11/08/2016 04:13 PM - Casey Bodley

<b>Status:</b>	Resolved	<b>Start date:</b>	11/08/2016
<b>Priority:</b>	High	<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>	other	<b>Reviewed:</b>	
<b>Tags:</b>		<b>Affected Versions:</b>	
<b>Backport:</b>	jewel	<b>ceph-qa-suite:</b>	
<b>Regression:</b>	No	<b>Pull request ID:</b>	
<b>Severity:</b>	3 - minor		

#### Description

The multisite test script test/rgw/test\_multi.py is getting stuck due to rejected requests. On the sending side:

```
2016-11-08 10:21:40.697972 7f3bb49f3700 20 cr:s=0x55f4a34ef570:op=0x55f4a363f500:17RGWFetchAllMeta
CR: operate()
2016-11-08 10:21:40.697992 7f3bb49f3700 20 cr:s=0x55f4a3a3aa30:op=0x55f4a358f000:13RGWomapAppend:
operate()
2016-11-08 10:21:40.698009 7f3bb49f3700 20 cr:s=0x55f4a34ef570:op=0x55f4a3640300:21RGWReadRESTReso
urceCRINSt7__cxx114listINS0_12basic_stringIcSt11char_traitsIcESaIcEEESaIS6_EEEE: operate()
2016-11-08 10:21:40.698081 7f3bb49f3700 20 RGWEnv::set(): HTTP_DATE: Tue Nov 8 15:21:40 2016
2016-11-08 10:21:40.698088 7f3bb49f3700 20 > HTTP_DATE -> Tue Nov 8 15:21:40 2016
2016-11-08 10:21:40.698096 7f3bb49f3700 10 get_canon_resource(): dest=/admin/metadata
2016-11-08 10:21:40.698097 7f3bb49f3700 10 generated canonical header: GET
2016-11-08 10:21:40.698138 7f3bb49f3700 15 generated auth header: AWS MGOSXQZS645INJSF:V3r0yc7VaXS
m49eKgo1Q+hcu4YA=
2016-11-08 10:21:40.698166 7f3bb49f3700 20 sending request to http://localhost:8000/admin/metadata
?rgwx-zonegroup=c4f11fe7-d7b7-4079-b7aa-4c4afcc978d4
2016-11-08 10:21:40.698179 7f3bb49f3700 20 register_request mgr=0x55f4a348e558 req_data->id=1, eas
y_handle=0x55f4a3a7a000
2016-11-08 10:21:40.698189 7f3bb49f3700 20 run: stack=0x55f4a34ef570 is io blocked
2016-11-08 10:21:40.698200 7f3bb51f4700 20 link_request req_data=0x55f4a3746640 req_data->id=1, ea
sy_handle=0x55f4a3a7a000
2016-11-08 10:21:40.698387 7f3bb51f4700 10 receive_http_header
2016-11-08 10:21:40.698393 7f3bb51f4700 10 received header:HTTP/1.1 400 Bad Request
2016-11-08 10:21:40.698424 7f3bb51f4700 10 receive_http_header
2016-11-08 10:21:40.698428 7f3bb51f4700 10 received header:Cache-Control: no-cache, no-store, must
-revalidate, private, max-age=0
2016-11-08 10:21:40.698436 7f3bb51f4700 10 receive_http_header
2016-11-08 10:21:40.698436 7f3bb51f4700 10 received header:Pragma: no-cache
2016-11-08 10:21:40.698439 7f3bb51f4700 10 receive_http_header
2016-11-08 10:21:40.698440 7f3bb51f4700 10 received header:Expires: 0
2016-11-08 10:21:40.698443 7f3bb51f4700 10 receive_http_header
2016-11-08 10:21:40.698445 7f3bb51f4700 10 received header>Date: Tue, 08 Nov 2016 15:21:40 GMT
2016-11-08 10:21:40.698449 7f3bb51f4700 10 receive_http_header
2016-11-08 10:21:40.698451 7f3bb51f4700 10 received header:Connection: close
2016-11-08 10:21:40.698454 7f3bb51f4700 10 receive_http_header
2016-11-08 10:21:40.698455 7f3bb51f4700 10 received header:
2016-11-08 10:21:40.698529 7f3bb49f3700 20 cr:s=0x55f4a34ef570:op=0x55f4a3640300:21RGWReadRESTReso
urceCRINSt7__cxx114listINS0_12basic_stringIcSt11char_traitsIcESaIcEEESaIS6_EEEE: operate()
2016-11-08 10:21:40.698605 7f3bb49f3700 20 cr:s=0x55f4a34ef570:op=0x55f4a3640300:21RGWReadRESTReso
urceCRINSt7__cxx114listINS0_12basic_stringIcSt11char_traitsIcESaIcEEESaIS6_EEEE: operate() returne
d r=-22
2016-11-08 10:21:40.698615 7f3bb49f3700 20 cr:s=0x55f4a34ef570:op=0x55f4a363f500:17RGWFetchAllMeta
CR: operate()
```

```
2016-11-08 10:21:40.698617 7f3bb49f3700 0 meta sync: ERROR: failed to fetch metadata sections
```

The receiving side's radosgw does not log anything about this request (no match for "admin/metadata").

After reverting the commit that updated civetweb to 1.8 (commit 3252c10a1523b9aa753fd4a88503eb5959b165f9), these requests succeed and the test\_multi.py completes successfully.

#### Related issues:

Copied to rgw - Backport #18112: jewel: multisite requests failing with '400 ...

**Resolved**

#### History

##### #1 - 11/08/2016 08:08 PM - Casey Bodley

i captured a failing request using tshark:

```
Hypertext Transfer Protocol
  GET /admin/log/?type=data&id=0&info&rgwx-zonegroup=16da5300-6a70-4a05-aa59-335ee0f815f7 HTTP/1.1\r\n
  [Expert Info (Chat/Sequence): GET /admin/log/?type=data&id=0&info&rgwx-zonegroup=16da5300-6a70-4a05-aa
59-335ee0f815f7 HTTP/1.1\r\n]
    [GET /admin/log/?type=data&id=0&info&rgwx-zonegroup=16da5300-6a70-4a05-aa59-335ee0f815f7 HTTP/1.1\
r\n]
      [Severity level: Chat]
      [Group: Sequence]
      Request Method: GET
      Request URI: /admin/log/?type=data&id=0&info&rgwx-zonegroup=16da5300-6a70-4a05-aa59-335ee0f815f7
      Request Version: HTTP/1.1
      Host: localhost:8000\r\n
      Accept: */*\r\n
      Transfer-Encoding: chunked\r\n
      AUTHORIZATION: AWS 0555b35654ad1656d804:BPIS02As9Smjzh5LGytJNtfUA44=\r\n
      DATE: Tue Nov 8 19:54:46 2016\r\n
      Expect: 100-continue\r\n
      \r\n
      [Full request URI: http://localhost:8000/admin/log/?type=data&id=0&info&rgwx-zonegroup=16da5300-6a70-4a05-
aa59-335ee0f815f7]
      [HTTP request 2/2]
      HTTP chunked response
        End of chunked encoding
        Chunk size: 0 octets
        trailer-part: Host: localhost:8000\r\nAccept: */*\r\nTransfer-Encoding: chunked\r\nAUTHORIZATION: AWS
0555b35654ad1656d804:BPIS02As9Smjzh5LGytJNtfUA44=\r\nDATE: Tue Nov 8 19:54:46 2016\r\nExpect: 100-continue\r\n
        \r\n
```

and the response:

```
Hypertext Transfer Protocol
  HTTP/1.1 400 Bad Request\r\n
  [Expert Info (Chat/Sequence): HTTP/1.1 400 Bad Request\r\n]
  [HTTP/1.1 400 Bad Request\r\n]
  [Severity level: Chat]
  [Group: Sequence]
  Request Version: HTTP/1.1
  Status Code: 400
  Response Phrase: Bad Request
  Cache-Control: no-cache, no-store, must-revalidate, private, max-age=0\r\n
  Pragma: no-cache\r\n
  Expires: 0\r\n
  Date: Tue, 08 Nov 2016 19:54:46 GMT\r\n
  Connection: close\r\n
  \r\n
  [HTTP response 2/2]
  [Time since request: 0.000112000 seconds]
  [Prev response in frame: 106]
  [Request in frame: 122]
```

## #2 - 11/08/2016 09:57 PM - Yehuda Sadeh

When sending an empty chunked request there still needs to be a marker that specifies end of data. Can you check to see if you see that? (the data portion of the request needs to include a 0)

## #3 - 11/08/2016 10:24 PM - Casey Bodley

Managed to catch the failure in `parse_http_message()`:

```
Breakpoint 4, parse_http_message (check_method=0, buf=0x55dd65d6e5f9 "Accept: /**\r\nTransfer-Encoding: chunked\r\nAUTHORIZATION: AWS 0555b35654ad1656d804:HhMgDGvqMREZCd7D5pyj9pKPrAc=\r\nDATE: Tue Nov 8 20:44:36 2016\r\nExpect: 100-continue\r\n\r", len=16384, ri=0x55dd65d6e000) at /home/cbodley/ceph/src/civetweb/src/civetweb.c:6667
6667         request_length = -1;
(gdb) list
6662         * "GET / HTTP/1.0 ...." or "HTTP/1.0 200 OK ..." */
6663         is_request =
6664             check_method ? is_valid_http_method(ri->request_method) : 1;
6665         if ((is_request && memcmp(ri->http_version, "HTTP/", 5) != 0)
6666             || (!is_request && memcmp(ri->request_method, "HTTP/", 5) != 0)) {
6667             request_length = -1;
6668         } else {
6669             if (is_request) {
6670                 ri->http_version += 5;
6671
(gdb) p ri->request_method
$1 = 0x55dd65d6e590 "/admin/metadata/user?rgwx-zonegroup=6dbd27c6-3191-4b2f-9134-2bd3518c36d9"
(gdb) p ri->request_uri
$2 = 0x55dd65d6e5d9 "HTTP/1.1\r\nHost:"
(gdb) p ri->http_version
$3 = 0x55dd65d6e5e9 "localhost:8000"
(gdb) up
#1 0x000055dd5c100008 in getreq (conn=0x55dd65d6e000, ebuf=0x7f0305ba7ab0 "", ebuf_len=100, err=0x7f0305ba7a8c)
    at /home/cbodley/ceph/src/civetweb/src/civetweb.c:11552
11552         } else if (parse_http_message(
(gdb) p conn->buf
$4 = 0x55dd65d6e590 "/admin/metadata/user?rgwx-zonegroup=6dbd27c6-3191-4b2f-9134-2bd3518c36d9"
```

This shows that `conn->buf` was missing the beginning of the request ("GET"), which led to parsing the wrong pieces into `request_method`, `request_uri` and `http_version`.

I looked through the git history for `src/civetweb`, and found commit `3aeb7a27b64fdae229b6ea9eff9fe2edb24e5739` "discard unread data if handled by `begin_request()`". Reverting that commit allows `test_multi.py` to run to completion, so we at least know it's related.

**#4 - 11/08/2016 10:46 PM - Yehuda Sadeh**

That's weird. Need to look at the previous request and what happens there. That commit fixes a similar issue, so I'm guessing that the discard fails when dealing with specific type of requests (chunked?).

**#5 - 11/17/2016 10:02 PM - Casey Bodley**

- File `http.multi.tcpdump` added

Attaching a raw tcpdump that captures the configuration (realm pull, period get, and period commit) and startup of two gateways on different zones. The master gateway is running on port 8000, which 'tcpdump -A -r http.multi.tcpdump' displays as 'localhost.irdmi'.

**#6 - 11/18/2016 10:39 PM - Yehuda Sadeh**

Seems that it's a chunked request without any data markers (need to specify a zero to end content).

**#7 - 11/18/2016 10:42 PM - Yehuda Sadeh**

What's also interesting is that it's a GET request with Transfer-Encoding header.

**#8 - 11/18/2016 10:45 PM - Yehuda Sadeh**

And now I also notice this:

```
GET /admin/log/?type=data&id=0&info&rgwx-zonegroup=7f0ed07b-95bf-4516-9389-f1640361986a HTTP/1.1
Host: localhost:8000
Accept: */*
Transfer-Encoding: chunked
AUTHORIZATION: AWS DiPt4V7WWvy2njL1z6aC:knm+EP6ApyAkWPFK2uWLFfX8E7g=
DATE: Thu Nov 17 21:52:16 2016
Expect: 100-continue
```

```
HTTP/1.1 200 OK
x-amz-request-id: tx00000000000000000000000008-00582e2690-100a-na-1
Content-Length: 38
Date: Thu, 17 Nov 2016 21:52:16 GMT
```

```
HTTP/1.1 200 OK
x-amz-request-id: tx00000000000000000000000008-00582e2690-100a-na-1
Content-Length: 38
Date: Thu, 17 Nov 2016 21:52:16 GMT
```

```
{"marker": "", "last_update": "0.000000"}
```

Instead of returning 100 response and then 200, there are two 200 responses.

**#9 - 11/18/2016 11:08 PM - Yehuda Sadeh**

just note that a single 200 response is also valid. Not sure why it would return two 200s, and I can't reproduce it locally.

**#10 - 11/21/2016 04:16 PM - Casey Bodley**

Thanks Yehuda. With 'tcpdump -A -r http.multi.tcpdump', I'm not seeing two different '200 OK' replies - I'm just seeing that reply split between two packets (one with headers, len=140, and one with the body, len=38).

The chunked GET requests do look wrong, and I think that's happening because we're setting CURLOPT\_UPLOAD=1 on all requests. I added a check to skip this when method == "GET", and that seems to fix things - multisite tests can run to completion.

**#11 - 11/21/2016 04:31 PM - Casey Bodley**

- Status changed from New to Need Review

- Backport set to jewel

<https://github.com/ceph/ceph/pull/12105>

**#12 - 12/01/2016 06:49 PM - Casey Bodley**

- Status changed from Need Review to Pending Backport

**#13 - 12/01/2016 09:18 PM - Nathan Cutler**

- Copied to Backport #18112: jewel: multisite requests failing with '400 Bad Request' with civetweb 1.8 added

**#14 - 12/08/2016 06:43 PM - Yehuda Sadeh**

- Status changed from Pending Backport to Resolved

**#15 - 02/01/2017 03:13 PM - Nathan Cutler**

Jewel backport was just merged, so now it's really resolved.

**Files**

---

http.multi.tcpdump	19.3 KB	11/17/2016	Casey Bodley
--------------------	---------	------------	--------------