

## Pulp - Issue #4603

### pulp\_streamer streams decodes responses, but sends the 'gzip' Content-Encoding header

03/29/2019 02:23 PM - dkliban@redhat.com

<b>Status:</b>	CLOSED - CURRENTRELEASE	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	dkliban@redhat.com	<b>Estimated time:</b>	0:00 hour
<b>Category:</b>		<b>Groomed:</b>	No
<b>Sprint/Milestone:</b>	2.19.1	<b>Sprint Candidate:</b>	No
<b>Severity:</b>	2. Medium	<b>Tags:</b>	Pulp 2
<b>Version:</b>		<b>Sprint:</b>	Sprint 51
<b>Platform Release:</b>	2.19.1	<b>Quarter:</b>	
<b>OS:</b>			
<b>Triaged:</b>	Yes		

#### Description

This does not occur 100% of the time, but often enough that it can be reproduced consistently.

- 1) Create an 'on\_demand' repository pointing to a RHEL 7 kickstart repository.
- 2) Sync the repository.
- 3) Download the repository

```
pulp-admin repo download --verify-all --repo-id ks
```

As the pulp\_worker was downloading the file the following was emitted by pulp\_streamer:

```
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] Unhandled Error
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] Traceback (most recent call last):
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] File "/usr/lib64/python2.7/site-packages
/twisted/application/app.py", line 399, in startReactor
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] self.config, oldstdout, oldstderr, sel
f.profiler, reactor)
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] File "/usr/lib64/python2.7/site-packages
/twisted/application/app.py", line 312, in runReactorWithLogging
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] reactor.run()
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] File "/usr/lib64/python2.7/site-packages
/twisted/internet/base.py", line 1261, in run
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] self.mainLoop()
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] File "/usr/lib64/python2.7/site-packages
/twisted/internet/base.py", line 1270, in mainLoop
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] self.runUntilCurrent()
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] --- <exception caught here> ---
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] File "/usr/lib64/python2.7/site-packages
/twisted/internet/base.py", line 869, in runUntilCurrent
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] f(*a, **kw)
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] File "/usr/lib64/python2.7/site-packages
/twisted/web/server.py", line 236, in write
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] http.Request.write(self, data)
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] File "/usr/lib64/python2.7/site-packages
/twisted/web/http.py", line 1110, in write
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] self.channel.writeSequence(toChunk(dat
a))
Mar 28 21:03:26 pulp2.dev pulp_streamer[17983]: [-] exceptions.AttributeError: 'NoneType' obje
ct has no attribute 'writeSequence'
```

Then the pulp\_worker emits the following:

```
Mar 28 21:03:26 pulp2.dev pulp[19562]: nectar.downloaders.threaded:ERROR: (19562-49088) ('Received
response with content-encoding: gzip, but failed to decode it.', error('Error -3 while decompress
ing: incorrect>
```

```
Mar 28 21:03:26 pulp2.dev pulp[19562]: nectar.downloaders.threaded:ERROR: (19562-49088) Traceback
(most recent call last):
Mar 28 21:03:26 pulp2.dev pulp[19562]: nectar.downloaders.threaded:ERROR: (19562-49088) File "/u
sr/lib/python2.7/site-packages/nectar/downloaders/threaded.py", line 292, in _fetch
Mar 28 21:03:26 pulp2.dev pulp[19562]: nectar.downloaders.threaded:ERROR: (19562-49088) for ch
unk in chunks:
Mar 28 21:03:26 pulp2.dev pulp[19562]: nectar.downloaders.threaded:ERROR: (19562-49088) File "/u
sr/lib/python2.7/site-packages/requests/models.py", line 755, in generate
Mar 28 21:03:26 pulp2.dev pulp[19562]: nectar.downloaders.threaded:ERROR: (19562-49088) raise
ContentDecodingError(e)
Mar 28 21:03:26 pulp2.dev pulp[19562]: nectar.downloaders.threaded:ERROR: (19562-49088) ContentDec
odingError: ('Received response with content-encoding: gzip, but failed to decode it.', error('Err
or -3 while de>
Mar 28 21:03:26 pulp2.dev pulp[19562]: pulp.server.controllers.repository:INFO: Download of /var/l
ib/pulp/content/units/distribution/ea/91381d46b89bcb5fd5f595b7bc0fc7802404b3ae4899901f8fb076af0c00
b2/images/pxeboot/initrd.img failed:
('Received response with content-encoding: gzip, but failed to decode it.', error('Error -3 while
decompressing: incorrect header check',)).
```

I was also able to cause this by simply requesting the initrd.img from the on\_demand repository.

This can also be reproduced with this repo[0].

[0] <https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-kickstart/>

#### Related issues:

Related to Pulp - Test #4628: Test pulp\_streamer stream decodes responses, bu...

**CLOSED - COMPLETE**

Related to Pulp - Issue #4649: Pulp 2 Nightly fails test\_package\_paths and te...

**CLOSED - WONTFIX**

#### Associated revisions

##### Revision 00a21c8d - 04/02/2019 10:28 PM - dkliban@redhat.com

Problem: downloader can't be configured to NOT decode

Solution: add downloader config option 'stream'

When 'stream' is True, the downloader does not decode the response. When 'stream' is False, the downloader decodes the response unless it's a file ending in .gz.

re: #4603 <https://pulp.plan.io/issues/4603>

##### Revision 48d365f1 - 04/02/2019 10:37 PM - dkliban@redhat.com

Problem: streamer decodes data while streaming

Solution: configure nectar downloaders to not decode responses

This patch relies on a new config option for the nectar downloaders.

re: #4603 <https://pulp.plan.io/issues/4603>

##### Revision f064a14c - 05/16/2019 07:19 PM - dkliban@redhat.com

Problem: streamer decodes data while streaming

Solution: configure nectar downloaders to not decode responses

This patch relies on a new config option for the nectar downloaders.

re: #4603 <https://pulp.plan.io/issues/4603> (cherry picked from commit 48d365f157e41b4b3ce4f0f33a0eab73ced8eea1)

##### Revision 7e1b98e9 - 05/28/2019 05:03 PM - bherring

Adding pulp\_streamer compression check tests

The problem occurs only for content that can be compressed with gzip. Nectar, the download library inside pulp workers, advertises that it can receive content compressed using gzip. When the worker asks the web server for one of the distribution files, the web server responds with the files compressed using gzip. The worker uses nectar to decode the file and then writes it to disk. The same nectar code is used in the streamer. When the streamer is passing along the file, it is passing it along in the decoded form. When the worker gets this data during the deferred\_download task or the download task, it is attempting to decompress it, but the streamer had already done that.

Adding a test to verify the content being published is compressed

closes #4603

## History

---

### #2 - 03/29/2019 03:20 PM - dkliban@redhat.com

- Status changed from NEW to ASSIGNED
- Assignee set to dkliban@redhat.com

### #3 - 03/29/2019 03:39 PM - CodeHeeler

- Triaged changed from No to Yes
- Sprint set to Sprint 51

### #4 - 03/30/2019 01:36 AM - dkliban@redhat.com

- Subject changed from *pulp\_streamer fails to properly stream response to pulp\_streamer streams decodes responses, but sends the 'gzip' Content-Encoding header*
- Description updated

The problem occurs only for content that can be compressed with gzip. nectar, the download library inside pulp workers, advertises that it can receive content compressed using gzip. When the worker asks the web server for one of the distribution files, the web server responds with the files compressed using gzip. The worker uses nectar to decode the file and then writes it to disk. The same nectar code is used in the streamer. So when the streamer is passing along the file, it is passing it along in the decoded form. So when the worker gets this data during the 'deferred\_download' task or the 'download' task, it is attempting to decompress it, but the streamer had already done that.

### #5 - 04/01/2019 11:03 PM - daviddavis

- Has duplicate Issue #4524: *Broken symlinks for subrepos are created during on\_demand sync of kickstart trees added*

### #6 - 04/02/2019 12:29 AM - dkliban@redhat.com

While python-requests allows the user to specify whether auto decoding should occur when reading the request data, nectar does not expose such functionality. Nectar's behaviour is hard-coded to NOT decode only files that have .gz extension[0]. In order to change that, the DownloaderConfig[1] needs to accept another option called 'stream'. When set to true, the HttpThreadedDownloader will stream raw bytes without decoding them. The streamer then needs to be updated to pass 'stream=True' to the DownloaderConfig that it creates for the downloaders.

[0] <https://github.com/pulp/nectar/blob/master/nectar/downloaders/threaded.py#L255>

[1] <https://github.com/pulp/nectar/blob/master/nectar/config.py#L9>

### #8 - 04/02/2019 02:52 PM - dkliban@redhat.com

- Status changed from ASSIGNED to NEW

### #9 - 04/02/2019 07:03 PM - dkliban@redhat.com

- Status changed from NEW to ASSIGNED

### #10 - 04/02/2019 07:15 PM - dkliban@redhat.com

The easiest way to reproduce the problem is to on\_demand sync a kickstart repo and then request an image file with 'Accept-Encoding' header set to 'gzip'. The response is supposed to be read by gzip, but as can be seen below, gzip doesn't recognize it.

```
pulp-admin rpm repo create --download-policy on_demand --feed https://repos.fedorapeople.org/pulp/pulp/fixtures/rpm-kickstart/ --repo-id ks
pulp-admin rpm repo sync run --repo-id ks
curl -sH 'Accept-encoding: gzip' -k -L https://localhost/pulp/repos/pulp/pulp/fixtures/rpm-kickstart/images/pxeboot/vmlinuz | gunzip -
```

```
gzip: stdin: not in gzip format
```

### #11 - 04/02/2019 07:28 PM - kersom

- Related to Test #4628: *Test pulp\_streamer stream decodes responses, but sends the 'gzip' Content-Encoding header added*

### #12 - 04/02/2019 07:55 PM - dkliban@redhat.com

- Status changed from ASSIGNED to POST

<https://github.com/pulp/nectar/pull/65>

<https://github.com/pulp/pulp/pull/3907>

**#13 - 04/09/2019 04:36 PM - dkliban@redhat.com**

- Related to Issue #4649: Pulp 2 Nightly fails test\_package\_paths and test\_download\_policies added

**#14 - 04/15/2019 12:21 AM - dkliban@redhat.com**

- Status changed from POST to MODIFIED

**#15 - 04/15/2019 09:57 PM - bmbouter**

- Tags Pulp 2 added

**#16 - 05/02/2019 02:54 PM - dkliban@redhat.com**

- Platform Release set to 2.19.1

**#17 - 05/03/2019 09:33 PM - dkliban@redhat.com**

- Sprint/Milestone set to 2.19.1

**#18 - 05/22/2019 09:26 PM - dkliban@redhat.com**

- Status changed from MODIFIED to 5

**#19 - 05/30/2019 06:42 PM - dkliban@redhat.com**

- Status changed from 5 to CLOSED - CURRENTRELEASE

**#20 - 07/25/2019 09:50 PM - dkliban@redhat.com**

- Has duplicate deleted (Issue #4524: Broken symlinks for subrepos are created during on\_demand sync of kickstart trees)