RPM Support - Issue #5951

yum_distributor wrongly skips publish after update of multi-repo errata

01/09/2020 06:02 AM - rmcgover

Status:	CLOSED - CURRENTRELEASE		Start date:	
Priority:	Normal		**Due date:**	
Assignee:	rmcgover		**Estimated time:**	0:00 hour
Category:				
Sprint/Milestone:				
Severity:	2. Medium		**Groomed:**	No
Version:			**Sprint Candidate:**	No
Platform Release:	2.21.1		**Tags:**	Pulp 2
OS:			**Sprint:**	
Triaged:	No		**Quarter:**	

Description

When multi-repo errata are used, publishing a repo may skip with "Repository content has not changed since last publish", even if erratum units in the repo have been changed since last publish.

Steps to reproduce

In summary: put an advisory into multiple repos, publish the repos, then update the advisory via any single repo and publish repos again. All repos other than the one used for the update will skip publish, even though the erratum unit has actually changed.

Detailed steps:

0. Ensure at least two RPM repos exist

e.g., assuming default dev-env setup:

```
pulp-admin rpm repo create --repo-id zoo2
```

1. Create new erratum in repo 1

Upload an RPM for the advisory:

```
pulp-admin rpm repo uploads rpm --repo-id zoo --file devel/pulp_rpm/plugins/test/data/walrus-5.21-
1.noarch.rpm
```

Then the advisory metadata:

```
echo walrus,5.21,1,0,noarch,walrus-5.21-1.noarch.rpm,e837a635cc99f967a70f34b268baa52e0f412c1502e08
e924ff5b09f1f9573f2,sha256,walrus-5.21-1.src.rpm > walrus-pkglist.csv
pulp-admin rpm repo uploads erratum --repo-id zoo --erratum-id RHBA-4020:1234 --title test --descr
iption test --version 1 --release 1 --type bugfix --status test --updated 2019-01-01 --issued test
 --from test --pkglist-csv walrus-pkglist.csv
```

2. Copy content into repo 2

```
pulp-admin rpm repo copy all --from-repo-id zoo --to-repo-id zoo2
```

3. Publish repos

```
pulp-admin rpm repo publish run --repo-id zoo
pulp-admin rpm repo publish run --repo-id zoo2
```

4. Verify erratum appears in repodata for both repos

Example:

```
$ sudo /bin/sh -c 'cd /var/lib/pulp/published/yum/master/yum_distributor/zoo/*/repodata; zcat $(eg
rep --only-matching "[^/]+updateinfo.xml.gz" ./repomd.xml)'
<?xml version="1.0" encoding="utf-8"?>
<updates>
  <update status="test" from="test" version="1" type="bugfix">
    <id>RHBA-4020:1234</id>
    <issued date="test" />
    <title>test</title>
    <release>1</release>
    <pushcount>1</pushcount>
    <description>test</description>
    <updated date="2019-01-01" />
    <references />
    <pkglist>
      <collection short="zoo_0_default">
        <name>zoo_0_default</name>
        <package src="walrus-5.21-1.src.rpm" name="walrus" epoch="0" version="5.21" release="1" ar
ch="noarch">
          <filename>walrus-5.21-1.noarch.rpm</filename>
          <sum type="sha256">e837a635cc99f967a70f34b268baa52e0f412c1502e08e924ff5b09f1f9573f2</sum
>
        </package>
      </collection>
    </pkglist>
  </update>
</updates>

$ sudo /bin/sh -c 'cd /var/lib/pulp/published/yum/master/yum_distributor/zoo2/*/repodata; zcat $(e
grep --only-matching "[^/]+updateinfo.xml.gz" ./repomd.xml)'
<?xml version="1.0" encoding="utf-8"?>
<updates>
  <update status="test" from="test" version="1" type="bugfix">
    <id>RHBA-4020:1234</id>
    <issued date="test" />
    <title>test</title>
    <release>1</release>
    <pushcount>1</pushcount>
    <description>test</description>
    <updated date="2019-01-01" />
    <references />
    <pkglist>
      <collection short="zoo2_0_default">
        <name>zoo2_0_default</name>
        <package src="walrus-5.21-1.src.rpm" name="walrus" epoch="0" version="5.21" release="1" ar
ch="noarch">
          <filename>walrus-5.21-1.noarch.rpm</filename>
          <sum type="sha256">e837a635cc99f967a70f34b268baa52e0f412c1502e08e924ff5b09f1f9573f2</sum
>
        </package>
      </collection>
    </pkglist>
  </update>
</updates>
```

5. Modify the erratum via repo 1

Here, description, version and updated are being changed.

```
pulp-admin rpm repo uploads erratum --repo-id zoo --erratum-id RHBA-4020:1234 --title test --descr
iption UPDATED --version 2 --release 1 --type bugfix --status test --updated 2020-01-01 --issued t
est --from test --pkglist-csv walrus-pkglist.csv
```

6. Publish the repos again

```
pulp-admin rpm repo publish run --repo-id zoo
pulp-admin rpm repo publish run --repo-id zoo2
```

7. Check repodata again

```
$ sudo /bin/sh -c 'cd /var/lib/pulp/published/yum/master/yum_distributor/zoo/*/repodata; zcat $(eg
rep --only-matching "[^/]+updateinfo.xml.gz" ./repomd.xml)'
<?xml version="1.0" encoding="utf-8"?>
<updates>
  <update status="test" from="test" version="2" type="bugfix">
    <id>RHBA-4020:1234</id>
    <issued date="test" />
    <title>test</title>
    <release>1</release>
    <pushcount>1</pushcount>
    <description>UPDATED</description>
    <updated date="2020-01-01" />
    <references />
    <pkglist>
      <collection short="zoo_0_default">
        <name>zoo_0_default</name>
        <package src="walrus-5.21-1.src.rpm" name="walrus" epoch="0" version="5.21" release="1" ar
ch="noarch">
          <filename>walrus-5.21-1.noarch.rpm</filename>
          <sum type="sha256">e837a635cc99f967a70f34b268baa52e0f412c1502e08e924ff5b09f1f9573f2</sum
>
        </package>
      </collection>
    </pkglist>
  </update>
</updates>

$ sudo /bin/sh -c 'cd /var/lib/pulp/published/yum/master/yum_distributor/zoo2/*/repodata; zcat $(e
grep --only-matching "[^/]+updateinfo.xml.gz" ./repomd.xml)'
<?xml version="1.0" encoding="utf-8"?>
<updates>
  <update status="test" from="test" version="1" type="bugfix">
    <id>RHBA-4020:1234</id>
    <issued date="test" />
    <title>test</title>
    <release>1</release>
    <pushcount>1</pushcount>
    <description>test</description>
    <updated date="2019-01-01" />
    <references />
    <pkglist>
      <collection short="zoo2_0_default">
        <name>zoo2_0_default</name>
        <package src="walrus-5.21-1.src.rpm" name="walrus" epoch="0" version="5.21" release="1" ar
ch="noarch">
          <filename>walrus-5.21-1.noarch.rpm</filename>
          <sum type="sha256">e837a635cc99f967a70f34b268baa52e0f412c1502e08e924ff5b09f1f9573f2</sum
>
        </package>
      </collection>
    </pkglist>
  </update>
</updates>
```

Actual behavior

updateinfo is only up to date in one of the published repos, even though publish was requested on both repos. Pulp skips publishing of the second repo.

Expected behavior

Neither publish is skipped, updateinfo is up to date in both published repos.

Additional info

For the upload use-case, it's possible for clients to work around this by explicitly importing the updated erratum into every repo containing the erratum (i.e. repeat step 5 for each repo containing the erratum). Although only the first import actually updates the units, the others will still reset last_unit_added on each repo, avoiding skipped publishes.

It seems possible that syncing would have the same problem (not tested).

Tested on current 2-master: bb9195f1dd30f51ed973024cbed8e087205749fc (pulp), cc35af04da2b9725db49290ebbc3253965951abf (pulp_rpm)

Associated revisions

Revision 94986211 - 01/22/2020 02:23 AM - rmcgover

Add controller method for multi-repo update of last_unit_added

In rare cases where mutating a single unit can affect multiple repos at once (e.g. erratum units), it's necessary to set last_unit_added on all relevant repos. Add a controller method for this, to be invoked by the relevant importer(s).

re: #5951 https://pulp.plan.io/issues/5951

Revision 69759d0f - 01/23/2020 04:10 AM - rmcgover

Ensure updating erratum sets last_unit_added on *all* repos

When we update an erratum, this affects not only the repo for which the update was requested, but also every other repo containing the same erratum.

Therefore, rather than letting the controller only set last_unit_added on the specific repo used for the import, we should ensure the field is set for all repos containing the unit - otherwise, publish of those repos may wrongly be skipped.

closes #5951 https://pulp.plan.io/issues/5951

Revision d73366bd - 02/03/2020 06:49 AM - rmcgover

Respect last_unit_added when deciding whether to publish

This is a follow-up to commits for issue #5951. Previous commits made it so that mutating an erratum unit would set last_unit_added on all affected repos; however, the code for deciding whether to skip publish ignored this field and only looked at RepositoryContentUnits, so the commits didn't fix the issue.

Fix this by checking last_unit_added as well, so we cover both the "new RepositoryContentUnits created for the repo" and "unit shared between multiple repos was updated" cases. Erratum units are the only type known to fall into the latter case.

re: #5951 https://pulp.plan.io/issues/5951

Revision 6fa51aab - 02/24/2020 11:42 AM - rmcgover

Ensure updating erratum sets last_unit_added on *all* repos

When we update an erratum, this affects not only the repo for which the update was requested, but also every other repo containing the same erratum.

Therefore, rather than letting the controller only set last_unit_added on the specific repo used for the import, we should ensure the field is set for all repos containing the unit - otherwise, publish of those repos may wrongly be skipped.

closes #5951 https://pulp.plan.io/issues/5951

(cherry picked from commit 69759d0fb9a16c0a47b1f49c78f6712e650912e1)

Revision 6b7272fd - 02/24/2020 11:45 AM - rmcgover

Add controller method for multi-repo update of last_unit_added

In rare cases where mutating a single unit can affect multiple repos at once (e.g. erratum units), it's necessary to set last_unit_added on all relevant repos. Add a controller method for this, to be invoked by the relevant importer(s).

re: #5951 https://pulp.plan.io/issues/5951 (cherry picked from commit 94986211526f637c31a6ff20655f7bbd72bbdd7b)

Revision 617236a4 - 02/24/2020 11:45 AM - rmcgover

Respect last_unit_added when deciding whether to publish

This is a follow-up to commits for issue #5951. Previous commits made it so that mutating an erratum unit would set last_unit_added on all affected repos; however, the code for deciding whether to skip publish ignored this field and only looked at RepositoryContentUnits, so the commits didn't fix the issue.

Fix this by checking last_unit_added as well, so we cover both the "new RepositoryContentUnits created for the repo" and "unit shared between multiple repos was updated" cases. Erratum units are the only type known to fall into the latter case.

re: #5951 https://pulp.plan.io/issues/5951 (cherry picked from commit d73366bda574e6f35b71ba274f00387e1cc22b85)

History

#1 - 01/10/2020 12:30 PM - ipanova@redhat.com

In Pulp2 errata is the only mutable content and publish does not check on that. The fix would be much more invasive than using the workaround and likely will have an impact on the performance. The workaround would be to trigger force_full publish. It will regenerate metadata from scratch and contain the updates performed on the errata. I do not foresee it fixed in Pulp2 being in maintenance mode.

There won't be such problem in Pulp 3.

#2 - 01/14/2020 06:45 AM - rmcgover

I'd like to point out that the workarounds probably are not as simple as it might seem at first glance:

For the force_full suggestion, we need to minimize the number of force_full publishes triggered because they can be extremely slow for large repos.

For the other workaround I mentioned in the report - upload the changed erratum to every repo containing it - the naive implementation would be "when updating an advisory, check the repos containing the advisory, then upload to each of them". However, that wouldn't be robust against being interrupted/cancelled and retried. If it were interrupted part way through the uploads and then retried, the advisory would already be up-to-date on the next attempt, so the needed uploads wouldn't be triggered. To be robust, the client will have to do unnecessary upload requests even when an advisory is already up-to-date or will have to add extra code to explicitly compare last_unit_added against the erratum timestamp on each repo.

As such, before committing to working around this in the clients I'd probably at least try a Pulp patch first.

#3 - 01/22/2020 02:36 AM - rmcgover

- Status changed from NEW to POST

- Assignee set to rmcgover

Pull requests for review: https://github.com/pulp/pulp/pull/3973, https://github.com/pulp/pulp_rpm/pull/1584.

#4 - 01/22/2020 10:23 AM - ppicka

- Tags Pulp 2 added

#5 - 01/23/2020 06:19 PM - rmcgover

- Status changed from POST to MODIFIED

Applied in changeset 69759d0fb9a16c0a47b1f49c78f6712e650912e1.

#6 - 02/18/2020 12:19 PM - ipanova@redhat.com

- Description updated

- Platform Release set to 2.21.1

#7 - 02/25/2020 02:33 PM - rmcgover

Applied in changeset 6fa51aabf719507759636761e2c4f33f54d741f9.

#8 - 02/27/2020 05:38 PM - ipanova@redhat.com

- Status changed from MODIFIED to 5

#9 - 03/04/2020 05:46 PM - ipanova@redhat.com

- Status changed from 5 to CLOSED - CURRENTRELEASE