RPM Support - Issue #9303

"xml:base" / "location_base" feature of RPM metadata is incompatible with some Pulp use cases, and is handled incorrectly in others

08/27/2021 09:59 AM - sjansen

Status:	CLOSED - CURRENTRELEASE	**Start date:**	
Priority:	High	**Due date:**	
Assignee:	dalley	**Estimated time:**	0:00 hour
Category:			
Sprint/Milestone:	3.16.0		
Severity:	2. Medium	**Groomed:**	No
Version:		**Sprint Candidate:**	No
Platform Release:		**Tags:**	Katello
OS:		**Sprint:**	Sprint 104
Triaged:	No	**Quarter:**	

Description

Environment: Katello 4.1.2 with pulp 3.14

python3-pulp-rpm-3.14.0-1.el8.noarch

python3-pulpcore-3.14.3-1.el8.noarch

katello-4.1.2.1-1.el8.noarch

How to reproduce:

1. Setup a repo. In this example i use harbottle as this repo as it contains an url in the location field in its metadata https://harbottle.gitlab.io/harbottle-main/8/x86_64/repodata/repomd.xml

2. Set Katello to "Mirror on Sync" and sync it

3. Run dnf update on client and get error as client tries to contact location urls outside of katello/pulp as metadata location field containts externel urls outside of pulp/katello

Snipped from https://harbottle.gitlab.io/harbottle-main/8/x86_64/repodata/repomd.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<repomd xmlns="http://linux.duke.edu/metadata/repo" xmlns:rpm="http://linux.duke.edu/metadata/rpm"
>
 <revision>1630020577</revision>
<data type="filelists">
  <checksum type="sha256">3acdbeee57c134225a95486260218f5234e81a4e6276f54697722c6dfd30d44f
</checksum>
  <open-checksum type="sha256">e914f59b5bd1cea374b3f82d68ab39b9a206d8ba46334cec5bdc049ed422d238
</open-checksum>
  <location xml:base="
https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64/" href="
repodata/3acdbeee57c134225a95486260218f5234e81a4e6276f54697722c6dfd30d44f-filelists.xml.gz"/>
  <timestamp>1630020612</timestamp>
  <size>200705</size>
  <open-size>2887999</open-size>
</data>
<data type="primary">
  <checksum type="sha256">388d9a1825cc25552147bc4445a4b696aff1f3c880ea04f7c870d7ee789716f8
</checksum>
  <open-checksum type="sha256">0522722ebad405d7c5e008592f343d75c35298fdb2c4d1af890ed3d44ce8eddd
</open-checksum>
  <location xml:base="
```

```
https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64/" href="
repodata/388d9a1825cc25552147bc4445a4b696aff1f3c880ea04f7c870d7ee789716f8-primary.xml.gz"/>
  <timestamp>1630020612</timestamp>
  <size>36580</size>
  <open-size>318118</open-size>
</data>~~~

**DNF Update:**

... el8_harbottle                                                    0.0  B/s |   0  B
00:00 Errors during downloading metadata for repository 'MyOrg_harbottle_el8_harbottle':
```

- Curl error (60): Peer certificate cannot be authenticated with given CA certificates for https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64/repodata/26425ae792fe92b77201ca2a024a13b2e158a681548d78a4916ed394b6f7a34c-filelists.xml.gz [SSL certificate problem: unable to get local issuer certificate]
- Curl error (60): Peer certificate cannot be authenticated with given CA certificates for https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64/repodata/cc45e1ec5af5a2f201c80dc7dc24284d50a09cb0c3f1542d239595ca3b86c0b6-primary.xml.gz [SSL certificate problem: unable to get local issuer certificate] Error: Failed to download metadata for repo 'MyOrg_harbottle_el8_harbottle': Yum repo downloading error: Downloading error(s): repodata/26425ae792fe92b77201ca2a024a13b2e158a681548d78a4916ed394b6f7a34c-filelists.xml.gz - Download failed: Curl error (60): Peer certificate cannot be authenticated with given CA certificates for https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64/repodata/26425ae792fe92b77201ca2a024a13b2e158a681548d78a4916ed394b6f7a34c-filelists.xml.gz [SSL certificate problem: unable to get local issuer certificate]; repodata/cc45e1ec5af5a2f201c80dc7dc24284d50a09cb0c3f1542d239595ca3b86c0b6-primary.xml.gz - Download failed: Curl error (60): Peer certificate cannot be authenticated with given CA certificates for https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64/repodata/cc45e1ec5af5a2f201c80dc7dc24284d50a09cb0c3f1542d239595ca3b86c0b6-primary.xml.gz [SSL certificate problem: unable to get local issuer certificate]

```
**Output of repomd.xml on katello/pulp (just on snipped):**

curl https://XXX/pulp/content/myOrg/Library/al8/custom/harbottle/el8_harbottle/repodata/repomd.xml

~~~ xml
<?xml version="1.0" encoding="UTF-8"?>
<repomd xmlns="http://linux.duke.edu/metadata/repo" xmlns:rpm="http://linux.duke.edu/metadata/rpm"
>
 <revision>1629937671</revision>
<data type="filelists">
  <checksum type="sha256">26425ae792fe92b77201ca2a024a13b2e158a681548d78a4916ed394b6f7a34c</checks
um>
  <open-checksum type="sha256">9fb8a946cd46d983b0ab1eb0bbac76b6a9069c63e07d193985385d9c52992d31</o
pen-checksum>
  <location xml:base="https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64
/" href="repodata/26425ae792fe92b77201ca2a024a13b2e158a681548d78a4916ed394b6f7a34c-filelists.xml.g
z"/>
  <timestamp>1629937709</timestamp>
  <size>200893</size>
  <open-size>2887999</open-size>
</data>
<data type="primary">
  <checksum type="sha256">cc45e1ec5af5a2f201c80dc7dc24284d50a09cb0c3f1542d239595ca3b86c0b6</checks
um>
  <open-checksum type="sha256">e4dfd3144903056448e689c303c77086c84d3b9515f0c7061642634555663102</o
pen-checksum>
  <location xml:base="https://copr-be.cloud.fedoraproject.org/results/harbottle/main/epel-8-x86_64
/" href="repodata/cc45e1ec5af5a2f201c80dc7dc24284d50a09cb0c3f1542d239595ca3b86c0b6-primary.xml.gz"
/>
  <timestamp>1629937709</timestamp>
  <size>36553</size>
  <open-size>318160</open-size>
</data>
```

One notice, this issue may disappear after another sync and comes back at some point. The only actions happ on this system is a daily sync, publish content view and deploy update by dnf update or remote exec. I try keep this system in this current state to help provide more information if required. I checked if upstream repo for metadata changes between this issues, but it always containts the externel references to copr.

Related issues:	
Related to RPM Support - Story #9316: As a user, I can mirror the packages in...	**CLOSED - CURRENTRELEASE**
Copied to RPM Support - Backport #9315: Backport #9303 "poor handling of xml:...	**CLOSED - CURRENTRELEASE**

Associated revisions

Revision b5744d6c - 08/31/2021 07:33 PM - dalley

Fix handling of location_base / xml:base

This metadata feature is incompatible with metadata mirroring. Reject repo syncs in this case and handle it more appropriately in other cases.

closes: #9303 https://pulp.plan.io/issues/9303

History

#1 - 08/30/2021 04:42 PM - dalley

- Status changed from NEW to ASSIGNED

- Priority changed from Normal to High

- Sprint set to Sprint 104

#2 - 08/30/2021 04:42 PM - dalley

- Project changed from Pulp to RPM Support

#3 - 08/30/2021 05:34 PM - dalley

- Assignee set to dalley

#4 - 08/30/2021 05:40 PM - dalley

- Status changed from ASSIGNED to NEW

- Assignee deleted (dalley)

#5 - 08/31/2021 12:30 AM - dalley

- Status changed from NEW to POST

- Assignee set to dalley

- Tags Katello added

The immediate cause of the error is a pretty simple fix, just take xml:base / location_base into consideration when generating the remote artifact records so that the download goes to the right place.

But after that it gets messy.

The goal of most Pulp users is to create a local clone of the repos, including the packages. "mirror" mode is intended to do as much as it can to keep repo the same, including keeping the same layout and making an exact copy of the metadata with intact metadata signatures.

But what this xml:base feature of the metadata does is basically incompatible with parts of that goal, because the metadata is telling the client to go look in some completely different place for the packages, probably over the internet.

So the best thing to do would be to just fail if you try to sync a repo that does this in "mirror" mode, and tell users to disable it for this repo. Thankfully it is not used very frequently.

However this does also mean that we probably need to re-introduce a way to mirror the packages in a repo (kick out ones that are no longer in the upstream) without actually mirroring the metadata.

#6 - 08/31/2021 12:32 AM - dalley

- Subject changed from pulp present "location field" unfiltered to client, maybe a security issue as externel urls are accessed by dnf update to "xml:base" / "location_base" feature of RPM metadata is incompatible with some Pulp use cases, and is handled incorrectly in others

#7 - 08/31/2021 12:40 AM - dalley

- Copied to Backport #9315: Backport #9303 "poor handling of xml:base / location_base" to 3.14.z added

#8 - 08/31/2021 12:53 AM - pulpbot

PR: https://github.com/pulp/pulp_rpm/pull/2110

#9 - 08/31/2021 05:46 AM - dalley

- Related to Story #9316: As a user, I can mirror the packages in a repo (kick out ones that are no longer in the upstream) without mirroring the metadata. added

#10 - 08/31/2021 07:33 PM - dalley

- Status changed from POST to MODIFIED

Applied in changeset b5744d6cdb592b644f269d5a5a930bfbc91d124b.

#11 - 10/20/2021 06:56 PM - pulpbot

- Sprint/Milestone set to 3.16.0

#12 - 10/20/2021 09:07 PM - pulpbot

- Status changed from MODIFIED to CLOSED - CURRENTRELEASE