

Katello - Bug #35332

Official SLES repos contain duplicate NEVRAs that can be synced to Katello server, but not to proxy

08/04/2022 01:38 PM - Quirin Pamp

Status:	Assigned	
Priority:	Normal	
Assignee:	Quirin Pamp	
Category:	Foreman Proxy Content	
Target version:	Katello Backlog	
Difficulty:		Fixed in Releases:
Triaged:	Yes	Found in Releases: Katello 4.1.4
Bugzilla link:		Red Hat JIRA:
Pull request:		

Description

Note: This issue strikes me as both a pulp_rpm issue as well as a "how Katello, SLES, and pulp_rpm interact" issue.

It looks like various SLES enterprise repos contain duplicate NEVRAs in the metadata which is obviously on SLES, but appears to work well enough for them. (Strictly speaking I believe it is undefined behaviour how packaging tools like yum or zypper handle such cases, I presume they just use the first duplicate NEVRA package they get to).

Example affected repo: `SLES12-SP3-Updates for sle-12-x86_64 in product SUSE Linux Enterprise Server 12 SP3 x86_64 (id: 1421)` (obtained via SCC manager plugin)

If one tries to sync such a repository to Katello using the "mirror_complete" mode, one gets the following sync error:

```
Parsing interrupted: The repository metadata being synced into Pulp is erroneous in a way that makes it ambiguous (duplicate NEVRAs), and therefore we do not allow it to be synced in 'mirror_complete' mode. Please choose a sync policy which does not mirror repository metadata.
```

Please read https://github.com/pulp/pulp_rpm/issues/2402 for more details.

see e.g. https://orcharhino.testdmz.atix/foreman_tasks/tasks/3ee0fe21-e380-48cc-8d97-41afb55e0016

At this point I can change the sync mode to "content_only" and sync the repo to Katello successfully.

However, if I then try to sync this repo to smart proxy, the smart proxy **always uses** "mirror_complete" mode and runs into the same error as before. As a user there is no way to tell the smart proxy sync to use anything other than "mirror_complete", so I am now stuck with a broken smart proxy sync.

Note that I found this issue on Katello 4.1, but given the pulp_rpm error has not changed in latest pulp_rpm, I believe this still affects the most recent Katello versions.

Internally we have "solved" this issue by patching pulp_rpm to downgrade the ERROR to a WARNING (though we have not yet done enough testing to be sure this does not cause issues when installing duplicate NEVRA packages on attached hosts)...

I have a few thoughts/questions aimed at Pulp RPM:

- Why are duplicate NEVRAs considered a bigger issue for "mirror_complete" mode, than for other sync policies? Shouldn't the opposite be true? If I am simply mirroring the upstream metadata, then I am passing through the issue from the upstream repo without taking responsibility for it. What the user gets should be as good or bad as the upstream repo itself, and if that worked for her, what Pulp serves should work just as well (or badly), no?
- If pulp lets us sync duplicate NEVRA repos using "content_only" mode, but what pulp then publishes still includes duplicate NEVRAs, and cannot be synced "mirror_complete" from Pulp to Pulp, then what have we gained by switching from "mirror_complete" to "content_only"?

I may be missing something (due to lack of in depth RPM repo knowledge), but the current handling of https://github.com/pulp/pulp_rpm/issues/2402 does not make any intuitive sense to me.

History

#1 - 08/04/2022 01:52 PM - Quirin Pamp

I also opened a pulp_rpm issue, so this can be discussed in both communities: https://github.com/pulp/pulp_rpm/issues/2691

#2 - 08/10/2022 05:44 PM - Chris Roberts

- Status changed from New to Need more information

Hi,

It looks like the pulp issue is merged, is this working for you now?

#3 - 08/11/2022 07:12 AM - Quirin Pamp

We still need to test the pulp_rpm patches, it is a complex test scenario requiring latest Pulp versions, but it is on my ToDo list.

#4 - 08/17/2022 05:41 PM - Chris Roberts

- Status changed from Need more information to Assigned

- Assignee set to Quirin Pamp

Moving to assigned based on feedback from Quirin

#5 - 08/31/2022 05:40 PM - Ian Ballou

- Category set to Foreman Proxy Content

- Target version set to Katello Backlog

- Triaged changed from No to Yes

Hey Quirin, we're going to set the version to the backlog for now, but feel free to unset that and unmark "triaged" once you get back to this issue.

#6 - 09/01/2022 07:48 AM - Quirin Pamp

@Ian, as far as we can tell this was fixed by the good people at pulp_rpm with the pulp_rpm 3.17.10 release.

I just saw this being packaged today: <https://github.com/theforeman/pulpcore-packaging/pull/558>

With that this should be fixed for Katello versions using pulpcore 3.18.

(We are still doing some testing around this whether this really fully fixes all the cases we were aware of, but I am pretty optimistic at this point.)

#7 - 09/01/2022 07:57 AM - Quirin Pamp

I consider this to be fixed. Also, it really was a pure pulp_rpm issue.

If we do still find some version of this problem with the fixed pulp_rpm versions, we will just open a new ticket with pulp_rpm.